(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 April 2002 (11.04.2002)

PCT

(10) International Publication Number
WO 02/29003 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/31243

(22) International Filing Date: 5 October 2001 (05.10.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/684,670 6 October 2000 (06.10.2000) US
60/300,894 26 June 2001 (26.06.2001) US

(71) Applicant and
(72) Inventor: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; West 116th Street and Broadway, New York, NY 10027 (US).

(72) Inventors: **JU, Jingyue**; 267 Marietta Street, Englewood Cliffs, NJ 07632 (US). **LI, Zengmin**; 647 West 172 Street, Apt. 41, New York, NY 10032 (US). **EDWARDS, John, Robert**; 485 Central Park West, #3C, New York, NY 10025 (US). **ITAGAKI, Yasuhiro**; 400 West 119th Street, #12D, New York, NY 10027 (US).

(74) Agent: **WHITE, John, P.**; Cooper & Dunahm LLP, 1185 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/29003 A2

(54) Title: MASSIVE PARALLEL METHOD FOR DECODING DNA AND RNA

(57) Abstract: This invention provides methods for attaching a nucleic acid to a solid surface and for sequencing nucleic acid by detecting the identity of each nucleotide analogue after the nucleotide analogue is incorporated into a growing strand of DNA in a polymerase reaction. The invention also provides nucleotide analogues which comprise unique labels attached to the nucleotide analogue through a cleavable linker, and a cleavable chemical group to cap the -OH group at the 3'-position of the deoxyribose.

MASSIVE PARALLEL METHOD FOR DECODING DNA AND RNA

This application claims the benefit of U.S. Provisional Application No. 60/300,894, filed June 26, 2001, and is a continuation-in-part of U.S. Serial No. 09/684,670, filed October 6, 2000, the contents of both of which are hereby incorporated by reference in their entireties into this application.

Background Of The Invention

Throughout this application, various publications are referenced in parentheses by author and year. Full citations for these references may be found at the end of the specification immediately preceding the claims. The disclosures of these publications in their entireties are hereby incorporated by reference into this application to more fully describe the state of the art to which this invention pertains.

The ability to sequence deoxyribonucleic acid (DNA) accurately and rapidly is revolutionizing biology and medicine. The confluence of the massive Human Genome Project is driving an exponential growth in the development of high throughput genetic analysis technologies. This rapid technological development involving chemistry, engineering, biology, and computer science makes it possible to move from studying single genes at a time to analyzing and comparing entire genomes.

heterozygotes unambiguously and are not 100% accurate in regions rich in nucleotides comprising guanine or cytosine due to compressions (Bowling et al. 1991; Yamakawa et al. 1997). In addition, the first few bases after the priming site are often masked by the high fluorescence signal from excess dye-labeled primers or dye-labeled terminators, and are therefore difficult to identify. Therefore, the requirement of electrophoresis for DNA sequencing is still the bottleneck for high-throughput DNA sequencing and mutation detection projects.

The concept of sequencing DNA by synthesis without using electrophoresis was first revealed in 1988 (Hyman, 1988) and involves detecting the identity of each nucleotide as it is incorporated into the growing strand of DNA in a polymerase reaction. Such a scheme coupled with the chip format and laser-induced fluorescent detection has the potential to markedly increase the throughput of DNA sequencing projects. Consequently, several groups have investigated such a system with an aim to construct an ultra high-throughput DNA sequencing procedure (Cheeseman 1994, Metzker et al. 1994). Thus far, no complete success of using such a system to unambiguously sequence DNA has been reported. The pyrosequencing approach that employs four natural nucleotides (comprising a base of adenine (A), cytosine (C), guanine (G), or thymine (T)) and several other enzymes for sequencing DNA by synthesis is now widely used for mutation detection (Ronaghi 1998). In this approach, the detection is based on the pyrophosphate (PPi) released during the DNA polymerase reaction, the

position of the deoxyribose ring in ddCTP is very crowded, while there is ample space for modification on the 5-position the cytidine base.

The approach disclosed in the present application is to make nucleotide analogues by linking a unique label such as a fluorescent dye or a mass tag through a cleavable linker to the nucleotide base or an analogue of the nucleotide base, such as to the 5-position of the pyrimidines (T and C) and to the 7-position of the purines (G and A), to use a small cleavable chemical moiety to cap the 3'-OH group of the deoxyribose to make it nonreactive, and to incorporate the nucleotide analogues into the growing DNA strand as terminators. Detection of the unique label will yield the sequence identity of the nucleotide. Upon removing the label and the 3'-OH capping group, the polymerase reaction will proceed to incorporate the next nucleotide analogue and detect the next base.

It is also desirable to use a photocleavable group to cap the 3'-OH group. However, a photocleavable group is generally bulky and thus the DNA polymerase will have difficulty to incorporate the nucleotide analogues containing a photocleavable moiety capping the 3'-OH group. If small chemical moieties that can be easily cleaved chemically with high yield can be used to cap the 3'-OH group, such nucleotide analogues should also be recognized as substrates for DNA polymerase. It has been reported that 3'-O-methoxy-deoxynucleotides are good substrates for several polymerases (Axelrod et al. 1978). 3'-O-allyl-dATP was also shown to be

the laser induced fluorescence detection approach which have overlapping fluorescence emission spectra, leading to heterozygote detection difficulty, the MS approach disclosed in this application produces very high resolution of sequencing data by detecting the cleaved small mass tags instead of the long DNA fragment. This method also produces extremely fast separation in the time scale of microseconds. The high resolution allows accurate digital mutation and heterozygote detection. Another advantage of sequencing with mass spectrometry by detecting the small mass tags is that the compressions associated with gel based systems are completely eliminated.

In order to maintain a continuous hybridized primer extension product with the template DNA, a primer that contains a stable loop to form an entity capable of self-priming in a polymerase reaction can be ligated to the 3' end of each single stranded DNA template that is immobilized on a solid surface such as a chip. This approach will solve the problem of washing off the growing extension products in each cycle.

Saxon and Bertozzi (2000) developed an elegant and highly specific coupling chemistry linking a specific group that contains a phosphine moiety to an azido group on the surface of a biological cell. In the present application, this coupling chemistry is adopted to create a solid surface which is coated with a covalently linked phosphine moiety, and to generate polymerase chain reaction (PCR) products that contain an azido group at the 5' end for specific coupling of the DNA

Summary Of The Invention

This invention is directed to a method for sequencing a nucleic acid by detecting the identity of a nucleotide analogue after the nucleotide analogue is incorporated into a growing strand of DNA in a polymerase reaction, which comprises the following steps:

(i) attaching a 5′ end of the nucleic acid to a solid surface;

(ii) attaching a primer to the nucleic acid attached to the solid surface;

(iii) adding a polymerase and one or more different nucleotide analogues to the nucleic acid to thereby incorporate a nucleotide analogue into the growing strand of DNA, wherein the incorporated nucleotide analogue terminates the polymerase reaction and wherein each different nucleotide analogue comprises (a) a base selected from the group consisting of adenine, guanine, cytosine, thymine, and uracil, and their analogues; (b) a unique label attached through a cleavable linker to the base or to an analogue of the base; (c) a deoxyribose; and (d) a cleavable chemical group to cap an -OH group at a 3′-position of the deoxyribose;

(iv) washing the solid surface to remove unincorporated nucleotide analogues;

wherein if the unique label is a dye, the order of steps (v) through (vii) is: (v), (vi), and (vii); and

wherein if the unique label is a mass tag, the order of steps (v) through (vii) is: (vi), (vii), and (v).

The invention provides a method of attaching a nucleic acid to a solid surface which comprises:

(i) coating the solid surface with a phosphine moiety,

(ii) attaching an azido group to a 5′ end of the nucleic acid, and

(iii) immobilizing the 5′ end of the nucleic acid to the solid surface through interaction between the phosphine moiety on the solid surface and the azido group on the 5′ end of the nucleic acid.

The invention provides a nucleotide analogue which comprises:

(a) a base selected from the group consisting of adenine or an analogue of adenine, cytosine or an analogue of cytosine, guanine or an analogue of guanine, thymine or an analogue of thymine, and uracil or an analogue of uracil;

Brief Description Of The Figures

Figure 1: The 3D structure of the ternary complexes of rat DNA polymerase, a DNA template-primer, and dideoxycytidine triphosphate (ddCTP). The left side of the illustration shows the mechanism for the addition of ddCTP and the right side of the illustration shows the active site of the polymerase. Note that the 3' position of the dideoxyribose ring is very crowded, while ample space is available at the 5 position of the cytidine base.

Figure 2A-2B: Scheme of sequencing by the synthesis approach. A: Example where the unique labels are dyes and the solid surface is a chip. B: Example where the unique labels are mass tags and the solid surface is channels etched into a glass chip. A, C, G, T; nucleotide triphosphates comprising bases adenine, cytosine, guanine, and thymine; d, deoxy; dd, dideoxy; R, cleavable chemical group used to cap the -OH group; Y, cleavable linker.

Figure 3: The synthetic scheme for the immobilization of an azido (N_3) labeled DNA fragment to a solid surface coated with a triarylphosphine moiety. Me, methyl group; P, phosphorus; Ph, phenyl.

Figure 4: The synthesis of triarylphosphine N-hydroxysuccinimide (NHS) ester.

Figure 5: The synthetic scheme for attaching an azido (N_3) group through a linker to the 5' end of a DNA

one, based on the complimentary template. The dye is detected and cleaved to test the approach. Dye1 = Fam; Dye2 = R6G; Dye3 = Tam; Dye4 = Rox.

Figure 10: The expected photocleavage products of DNA containing a photo-cleavable dye (Tam). Light absorption (300 - 360 nm) by the aromatic 2-nitrobenzyl moiety causes reduction of the 2-nitro group to a nitroso group and an oxygen insertion into the carbon-hydrogen bond located in the 2-position followed by cleavage and decarboxylation (Pillai 1980).

Figure 11: Synthesis of **PC-LC-Biotin-FAM** to evaluate the photolysis efficiency of the fluorophore coupled with the photocleavable linker 2-nitrobenzyl group.

Figure 12: Fluorescence spectra (λ_{ex} = 480 nm) of **PC-LC-Biotin-FAM** immobilized on a microscope glass slide coated with streptavidin (**a**); after 10 min photolysis (λ_{irr} = 350 nm; ~0.5 mW/cm^2) (**b**); and after washing with water to remove the photocleaved dye (**c**).

Figure 13A-13B: Synthetic scheme for capping the 3′-OH of nucleotide.

Figure 14: Chemical cleavage of the MOM group (top row) and the allyl group (bottom row) to free the 3′-OH in the nucleotide. CITMS = chlorotrimethylsilane.

Figure 15A-15B: Examples of energy transfer coupled dye systems, where Fam or Cy2 is employed as a light absorber (energy transfer donor) and Cl_2Fam, Cl_2R6G,

Figure 20: Example of synthesis of NHS ester of one mass tag (Tag-3). A similar scheme is used to create other mass tags.

Figure 21: A representative scheme for the synthesis of the nucleotide analogue $_{3'\text{-RO}}\text{-G-}_{\text{Tag3}}$. A similar scheme is used to create the other three modified bases $_{3'\text{-RO}}\text{-A-}_{\text{Tag1}}$, $_{3'\text{-RO}}\text{-C-}_{\text{Tag2}}$, $_{3'\text{-RO}}\text{-T-}_{\text{Tag4}}$. (i) tetrakis(triphenylphosphine)palladium(0); (ii) $POCl_3$, Bn_4N^+pyrophosphate; (iii) NH_4OH; (iv) $Na_2CO_3/NaHCO_3$ (pH = 9.0)/DMSO.

Figure 22: Examples of expected photocleavage products of DNA containing a photocleavable mass tag.

Figure 23: System for DNA sequencing comprising multiple channels in parallel and multiple mass spectrometers in parallel. The example shows 96 channels in a silica glass chip.

Figure 24: Parallel mass spectrometry system for DNA sequencing. Example shows three mass spectrometers in parallel. Samples are injected into the ion source where they are mixed with a nebulizer gas and ionized. A turbo pump is used to continuously sweep away free radicals, neutral compounds and other undesirable elements coming from the ion source. A second turbo pump is used to generate a continuous vacuum in all three analyzers and detectors simultaneously. The acquired signal is then converted to a digital signal by the A/D converter. All three signals are then sent to

Detailed Description Of The Invention

The following definitions are presented as an aid in understanding this invention.

As used herein, to **cap** an -OH group means to replace the "H" in the -OH group with a chemical group. As disclosed herein, the -OH group of the nucleotide analogue is capped with a cleavable chemical group. To **uncap** an -OH group means to cleave the chemical group from a capped -OH group and to replace the chemical group with "H", i.e., to replace the "R" in -OR with "H" wherein "R" is the chemical group used to cap the -OH group.

The nucleotide bases are abbreviated as follows: adenine (A), cytosine (C), guanine (G), thymine (T), and uracil (U).

An **analogue** of a **nucleotide base** refers to a structural and functional derivative of the base of a nucleotide which can be recognized by polymerase as a substrate. That is, for example, an analogue of adenine (A) should form hydrogen bonds with thymine (T), a C analogue should form hydrogen bonds with G, a G analogue should form hydrogen bonds with C, and a T analogue should form hydrogen bonds with A, in a double helix format. Examples of analogues of nucleotide bases include, but are not limited to, 7-deaza-adenine and 7-deaza-guanine, wherein the nitrogen atom at the 7-position of adenine or guanine is substituted with a carbon atom.

The present invention is directed to a method for sequencing a nucleic acid by detecting the identity of a nucleotide analogue after the nucleotide analogue is incorporated into a growing strand of DNA in a polymerase reaction, which comprises the following steps:

(i) attaching a 5′ end of the nucleic acid to a solid surface;

(ii) attaching a primer to the nucleic acid attached to the solid surface;

(iii) adding a polymerase and one or more different nucleotide analogues to the nucleic acid to thereby incorporate a nucleotide analogue into the growing strand of DNA, wherein the incorporated nucleotide analogue terminates the polymerase reaction and wherein each different nucleotide analogue comprises (a) a base selected from the group consisting of adenine, guanine, cytosine, thymine, and uracil, and their analogues; (b) a unique label attached through a cleavable linker to the base or to an analogue of the base; (c) a deoxyribose; and (d) a cleavable chemical group to cap an -OH group at a 3′-position of the deoxyribose;

(iv) washing the solid surface to remove unincorporated nucleotide analogues;

wherein if the unique label is a mass tag, the order of steps (v) through (vii) is: (vi), (vii), and (v).

In one embodiment of any of the nucleotide analogues described herein, the nucleotide base is adenine. In one embodiment, the nucleotide base is guanine. In one embodiment, the nucleotide base is cytosine. In one embodiment, the nucleotide base is thymine. In one embodiment, the nucleotide base is uracil. In one embodiment, the nucleotide base is an analogue of adenine. In one embodiment, the nucleotide base is an analogue of guanine. In one embodiment, the nucleotide base is an analogue of cytosine. In one embodiment, the nucleotide base is an analogue of thymine. In one embodiment, the nucleotide base is an analogue of uracil.

In different embodiments of any of the inventions described herein, the solid surface is glass, silicon, or gold. In different embodiments, the solid surface is a magnetic bead, a chip, a channel in a chip, or a porous channel in a chip. In one embodiment, the solid surface is glass. In one embodiment, the solid surface is silicon. In one embodiment, the solid surface is gold. In one embodiments, the solid surface is a magnetic bead. In one embodiment, the solid surface is a chip. In one embodiment, the solid surface is a channel in a chip. In one embodiment, the solid surface is a porous channel in a chip. Other materials can also be used as long as the material does not interfere with the steps of the method.

the solid surface is a ribonucleic acid (RNA), and the polymerase in step (iii) is reverse transcriptase.

In one embodiment, the primer is attached to a 3′ end of the nucleic acid in step (ii), and the attached primer comprises a stable loop and an -OH group at a 3′-position of a deoxyribose capable of self-priming in the polymerase reaction. In one embodiment, the step of attaching the primer to the nucleic acid comprises hybridizing the primer to the nucleic acid or ligating the primer to the nucleic acid. In one embodiment, the primer is attached to the nucleic acid through a ligation reaction which links the 3′ end of the nucleic acid with the 5′ end of the primer.

In one embodiment, one or more of four different nucleotide analogs is added in step (iii), wherein each different nucleotide analogue comprises a different base selected from the group consisting of thymine or uracil or an analogue of thymine or uracil, adenine or an analogue of adenine, cytosine or an analogue of cytosine, and guanine or an analogue of guanine, and wherein each of the four different nucleotide analogues comprises a unique label.

In one embodiment, the cleavable chemical group that caps the -OH group at the 3′-position of the deoxyribose in the nucleotide analogue is $-CH_2OCH_3$ or $-CH_2CH=CH_2$. Any chemical group could be used as long as the group 1) is stable during the polymerase reaction, 2) does not interfere with the recognition of the nucleotide

In one embodiment, the unique label that is attached to the nucleotide analogue is a mass tag that can be detected and differentiated by a mass spectrometer. In further embodiments, the mass tag is selected from the group consisting of a 2-nitro-α-methyl-benzyl group, a 2-nitro-α-methyl-3-fluorobenzyl group, a 2-nitro-α-methyl-3,4-difluorobenzyl group, and a 2-nitro-α-methyl-3,4-dimethoxybenzyl group. In one embodiment, the mass tag is a 2-nitro-α-methyl-benzyl group. In one embodiment, the mass tag is a 2-nitro-α-methyl-3-fluorobenzyl group. In one embodiment, the mass tag is a 2-nitro-α-methyl-3,4-difluorobenzyl group. In one embodiment, the mass tag is a 2-nitro-α-methyl-3,4-dimethoxybenzyl group. In one embodiment, the mass tag is detected using a parallel mass spectrometry system which comprises a plurality of atmospheric pressure chemical ionization mass spectrometers for parallel analysis of a plurality of samples comprising mass tags.

In one embodiment, the unique label is attached through a cleavable linker to a 5-position of cytosine or thymine or to a 7-position of deaza-adenine or deaza-guanine. The unique label could also be attached through a cleavable linker to another position in the nucleotide analogue as long as the attachment of the label is stable during the polymerase reaction and the nucleotide analog can be recognized by polymerase as a substrate. For example, the cleavable label could be attached to the deoxyribose.

different dideoxynucleotides are selected from the group consisting of 2',3'-dideoxyadenosine 5'-triphosphate, 2',3'-dideoxyguanosine 5'-triphosphate, 2',3'-dideoxycytidine 5'-triphosphate, 2',3'-dideoxythymidine 5'-triphosphate, 2',3'-dideoxyuridine 5'-triphosphase, and their analogues. In one embodiment, the dideoxynucleotide is 2',3'-dideoxyadenosine 5'-triphosphate. In one embodiment, the dideoxynucleotide is 2',3'-dideoxyguanosine 5'-triphosphate. In one embodiment, the dideoxynucleotide is 2',3'-dideoxycytidine 5'-triphosphate. In one embodiment, the dideoxynucleotide is 2',3'-dideoxythymidine 5'-triphosphate. In one embodiment, the dideoxynucleotide is 2',3'-dideoxyuridine 5'-triphosphase. In one embodiment, the dideoxynucleotide is an analogue of 2',3'-dideoxyadenosine 5'-triphosphate. In one embodiment, the dideoxynucleotide is an analogue of 2',3'-dideoxyguanosine 5'-triphosphate. In one embodiment, the dideoxynucleotide is an analogue of 2',3'-dideoxycytidine 5'-triphosphate. In one embodiment, the dideoxynucleotide is an analogue of 2',3'-dideoxythymidine 5'-triphosphate. In one embodiment, the dideoxynucleotide is an analogue of 2',3'-dideoxyuridine 5'-triphosphase.

In one embodiment, a polymerase and one or more of four different dideoxynucleotides are added in step (vi), wherein each different dideoxynucleotide is selected from the group consisting of 2',3'-dideoxyadenosine 5'-triphosphate or an analogue of 2',3'-dideoxyadenosine 5'-triphosphate; 2',3'-dideoxyguanosine 5'-triphosphate or an analogue of 2',3'-dideoxyguanosine 5'-

The invention provides a method for simultaneously sequencing a plurality of different nucleic acids, which comprises simultaneously applying any of the methods disclosed herein for sequencing a nucleic acid to the plurality of different nucleic acids. In different embodiments, the method can be used to sequence from one to over 100,000 different nucleic acids simultaneously.

The invention provides for the use of any of the methods disclosed herein for detection of single nucleotide polymorphisms, genetic mutation analysis, serial analysis of gene expression, gene expression analysis, identification in forensics, genetic disease association studies, DNA sequencing, genomic sequencing, translational analysis, or transcriptional analysis.

The invention provides a method of attaching a nucleic acid to a solid surface which comprises:

(i) coating the solid surface with a phosphine moiety,

(ii) attaching an azido group to a 5′ end of the nucleic acid, and

(iii) immobilizing the 5′ end of the nucleic acid to the solid surface through interaction between the phosphine moiety on the solid surface and the azido group on the 5′ end of the nucleic acid.

(a) a base selected from the group consisting of adenine or an analogue of adenine, cytosine or an analogue of cytosine, guanine or an analogue of guanine, thymine or an analogue of thymine, and uracil or an analogue of uracil;

(b) a unique label attached through a cleavable linker to the base or to an analogue of the base;

(c) a deoxyribose; and

(d) a cleavable chemical group to cap an -OH group at a 3'-position of the deoxyribose.

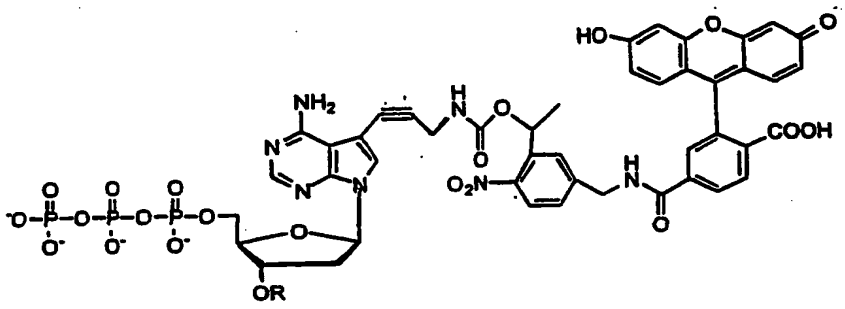
In one embodiment of the nucleotide analogue, the cleavable chemical group that caps the -OH group at the 3'-position of the deoxyribose is $-CH_2OCH_3$ or $-CH_2CH=CH_2$.

In one embodiment, the unique label is a fluorescent moiety or a fluorescent semiconductor crystal. In further embodiments, the fluorescent moiety is selected from the group consisting of 5-carboxyfluorescein, 6-carboxyrhodamine-6G, N,N,N',N'-tetramethyl-6-carboxyrhodamine, and 6-carboxy-X-rhodamine.

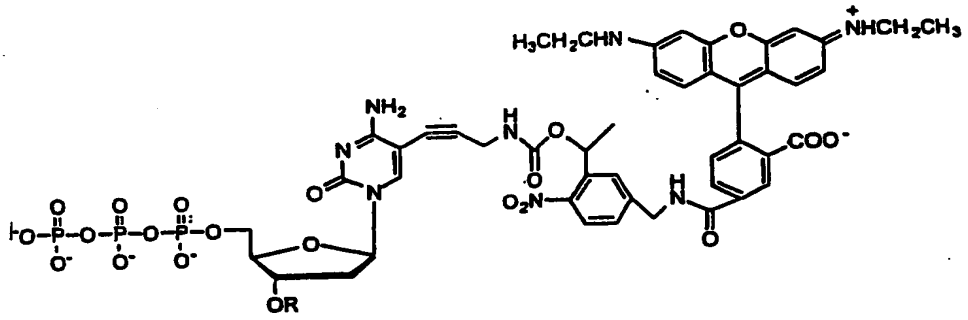
In one embodiment, the unique label is a fluorescence energy transfer tag which comprises an energy transfer donor and an energy transfer acceptor. In further embodiments, the energy transfer donor is 5-carboxyfluorescein or cyanine, and wherein the energy

In one embodiment, the cleavable chemical group used to cap the -OH group at the 3'-position of the deoxyribose is cleavable by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

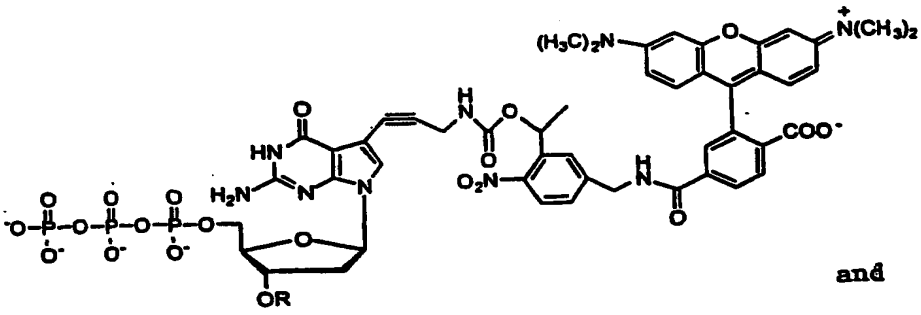
In different embodiments, the nucleotide analogue is selected from the group consisting of:



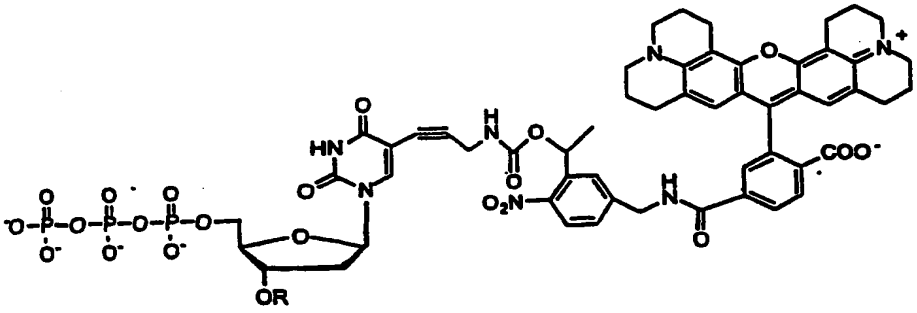
,



,



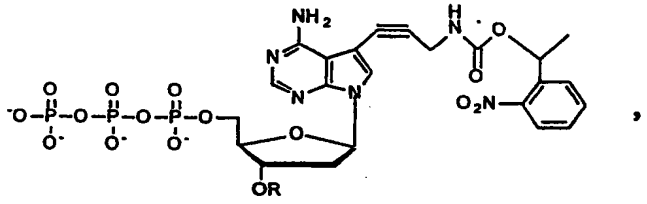
and

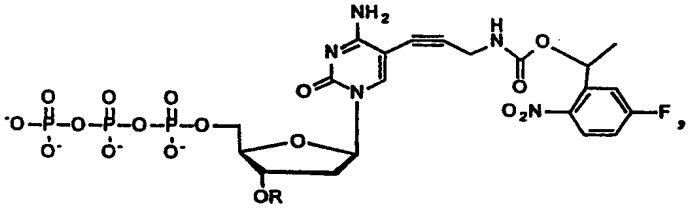


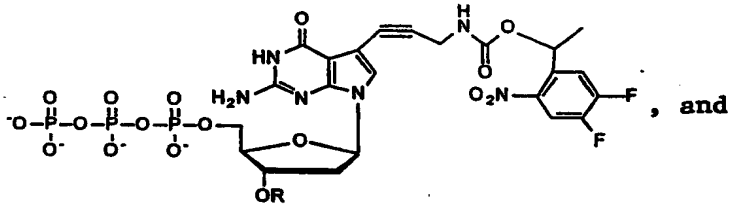
;

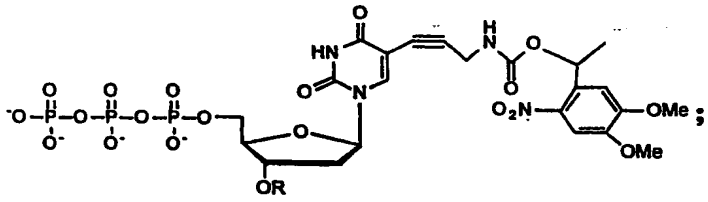
wherein R is $-CH_2OCH_3$ or $-CH_2CH=CH_2$.

In different embodiments, the nucleotide analogue is selected from the group consisting of:


,


,


, and


;

wherein R is $-CH_2OCH_3$ or $-CH_2CH=CH_2$.

This invention will be better understood from the Experimental Details which follow. However, one skilled in the art will readily appreciate that the specific methods and results discussed are merely illustrative of the invention as described more fully in the claims which follow thereafter.

(DMSO)/$NaHCO_3$ (pH=8.2) overnight at room temperature produces **PC-LC-Biotin-FAM** which is composed of a biotin at one end, a photocleavable 2-nitrobenzyl group in the middle, and a dye tag (FAM) at the other end. This photocleavable moiety closely mimics the designed photocleavable nucleotide analogues shown in **Figure 10**. Thus the successful photolysis of the **PC-LC-Biotin-FAM** moiety provides proof of the principle of high efficiency photolysis as used in the DNA sequencing system. For photolysis study, **PC-LC-Biotin-FAM** is first immobilized on a microscope glass slide coated with streptavidin (XENOPORE, Hawthorne NJ). After washing off the non-immobilized **PC-LC-Biotin-FAM**, the fluorescence emission spectrum of the immobilized **PC-LC-Biotin-FAM** was taken as shown in **Figure 12** (Spectrum **a**). The strong fluorescence emission indicates that **PC-LC-Biotin-FAM** is successfully immobilized to the streptavidin coated slide surface. The photocleavability of the 2-nitrobenzyl linker by irradiation at 350 nm was then tested. After 10 minutes of photolysis (λ_{irr} = 350 nm; ~0.5 mW/cm^2) and before any washing, the fluorescence emission spectrum of the same spot on the slide was taken that showed no decrease in intensity (**Figure 12**, Spectrum **b**), indicating that the dye (FAM) was not bleached during the photolysis process at 350 nm. After washing the glass slide with HPLC water following photolysis, the fluorescence emission spectrum of the same spot on the slide showed significant intensity decrease (**Figure 12**, Spectrum **c**) which indicates that most of the fluorescence dye (FAM) was cleaved from the immobilized biotin moiety and was removed by the washing procedure. This experiment shows

The ET primer and ET dideoxynucleotides have been shown to be a superior set of reagents for 4-color DNA sequencing that allows the use of one laser to excite multiple sets of fluorescent tags (Ju et al. 1995). It has been shown that DNA polymerase (Thermo Sequenase and Taq FS) can efficiently incorporate the ET dye labeled dideoxynucleotides (Rosenblum et al. 1997). These ET dye-labeled sequencing reagents are now widely used in large scale DNA sequencing projects, such as the human genome project. A library of ET dye labeled nucleotide analogues can be synthesized as shown in **Figure 15** for optimization of the DNA sequencing system. The ET dye set (FAM-Cl_2FAM, FAM-Cl_2R6G, FAM-Cl_2TAM, FAM-Cl_2ROX) using FAM as a donor and dichloro(FAM, R6G, TAM, ROX) as acceptors has been reported in the literature (Lee et al. 1997) and constitutes a set of commercially available DNA sequencing reagents. These ET dye sets have been proven to produce enhanced fluorescence intensity, and the nucleotides labeled with these ET dyes at the 5-position of T and C and the 7-position of G and A are excellent substrates of DNA polymerase. Alternatively, an ET dye set can be constructed using cyanine (Cy2) as a donor and Cl_2FAM, Cl_2R6G, Cl_2TAM, or Cl_2ROX as energy acceptors. Since Cy2 possesses higher molar absorbance compared with the rhodamine and fluorescein derivatives, an ET system using Cy2 as a donor produces much stronger fluorescence signals than the system using FAM as a donor (Hung et al. 1996). **Figure 16** shows a synthetic scheme for an ET dye labeled nucleotide analogue with Cy2 as a donor and Cl_2FAM as an acceptor using similar coupling chemistry as for the synthesis of an energy transfer system using FAM as a

and Tag4 are four different unique cleavable mass tags. Four specific examples of nucleotide analogues are shown in **Figure 19**. In **Figure 19**, "R" is H when the 3'-OH group is not capped. As discussed above, the photo cleavable 2-nitro benzyl moiety has been used to link biotin to DNA and protein for efficient removal by UV light (~ 350 nm) irradiation (Olejnik et al. 1995, 1999). Four different 2-nitro benzyl groups with different molecular weights as mass tags are used to form the mass tag labeled nucleotides as shown in **Figure 19**: 2-nitro-α-methyl-benzyl (Tag-1) codes for A; 2-nitro-α-methyl-3-fluorobenzyl (Tag-2) codes for C; 2-nitro-α-methyl-3,4-difluorobenzyl (Tag-3) codes for G; 2-nitro-α-methyl-3,4-dimethoxybenzyl (Tag-4) codes for T.

As a representative example, the synthesis of the NHS ester of one mass tag (Tag-3) is shown in **Figure 20**. A similar scheme is used to create the other mass tags. The synthesis of $_{3'\text{-HO}}\text{-G-}_{\text{Tag3}}$ is shown in **Figure 21** using well-established procedures (Prober et al. 1987; Lee et al. 1992 and Hobbs et al. 1991). 7-propargylamino- dGTP is first prepared by reacting 7-I-dGTP with N-trifluoroacetylpropargyl amine, which is then coupled with the NHS-Tag-3 to produce $_{3'\text{-HO}}\text{-G-}_{\text{Tag3}}$. The nucleotide analogues with a free 3'-OH are good substrates for the polymerase.

The sequencing by synthesis approach can be tested using mass tags using a scheme similar to that show for dyes in **Figure 9**. A DNA template containing a portion of nucleotide sequence that has no repeated sequences after

chemically with high yield as shown in **Figure 14** (Ireland, et al. 1986; Kamal et al. 1999). The chemical cleavage of the MOM and allyl groups is fairly mild and specific, so as not to degrade the DNA template moiety.

7. Parallel Channel System for Sequencing by Synthesis

Figure 23 illustrates an example of a parallel channel system. The system can be used with mass tag labels as shown and also with dye labels. A plurality of channels in a silica glass chip are connected on each end of the channel to a well in a well plate. In the example shown there are 96 channels each connected to its own wells. The sequencing system also permits a number of channels other than 96 to be used. 96 channel devices for separating DNA sequencing and sizing fragments have been reported (Woolley and Mathies 1994, Woolley et al. 1997, Simpson et al. 1998). The chip is made by photolithographic masking and chemical etching techniques. The photolithographically defined channel patterns are etched in a silica glass substrate, and then capillary channels (id ~ 100 μm) are formed by thermally bonding the etched substrate to a second silica glass slide. Channels are porous to increase surface area. The immobilized single stranded DNA template chip is prepared according to the scheme shown in **Figure 3.** Each channel is first treated with 0.5 M NaOH, washed with water, and is then coated with high density 3-aminopropyltrimethoxysilane in aqueous ethanol (Woolley et al. 1994) forming a primary amine surface. Succinimidyl (NHS) ester of triarylphosphine (**1**) is covalently coupled with the primary amine group converting the amine surface to a novel triarylphosphine

To make mass spectrometry competitive with a 96 capillary array method for analyzing DNA, a parallel mass spectrometer approach is needed. Such a complete system has not been reported mainly due to the fact that most of the mass spectrometers are designed to achieve adequate resolution for large biomolecules. The system disclosed herein requires the detection of four mass tags, with molecular weight range between 150 and 250 daltons, coding for the identity of the four nucleotides (A, C, G, T) . Since a mass spectrometer dedicated to detection of these mass tags only requires high resolution for the mass range of 150 to 250 daltons instead of covering a wide mass range, the mass spectrometer can be miniaturized and have a simple design. Either quadrupole (including ion trap detector) or time-of-flight mass spectrometers can be selected for the ion optics. While modern mass spectrometer technology has made it possible to produce miniaturized mass spectrometers, most current research has focused on the design of a single stand-alone miniaturized mass spectrometer. Individual components of the mass spectrometer has been miniaturized for enhancing the mass spectrometer analysis capability (Liu et al. 2000, Zhang et al. 1999). A miniaturized mass spectrometry system using multiple analyzers (up to 10) in parallel has been reported (Badman and Cooks 2000). However, the mass spectrometer of Badman and Cook was designed to measure only single samples rather than multiple samples in parallel. They also noted that the miniaturization of the ion trap limited the capability of the mass spectrometer to scan wide mass ranges. Since the

scanning mode of mass spectrometers are the same for each miniaturized mass spectrometer, one power supply for each analyzer and the ionization source can provide the necessary power for all three instruments. One power supply for each of the three independent detectors is used for spectrum collection. The data obtained are transferred to three independent A/D converters and processed by the data system simultaneously to identify the mass tag in the injected sample and thus identify the nucleotide. Despite containing three mass spectrometers, the entire device is able to fit on a laboratory bench top.

9. Validate the Complete Sequencing by Synthesis System By Sequencing P53 Genes

The tumor suppressor gene p53 can be used as a model system to validate the DNA sequencing system. The p53 gene is one of the most frequently mutated genes in human cancer (O'Connor et al. 1997). First, a base pair DNA template (shown below) is synthesized containing an azido group at the 5' end and a portion of the sequences from exon 7 and exon 8 of the p53 gene:

5'-N_3-TTCCTGCATGGGCGGCATGAACCCGAGGCCCATCCTCACCATCATCAC
ACTGGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTTGAGGTGCATT
-3' (SEQ ID NO: 2).

This template is chosen to explore the use of the sequencing system for the detection of clustered hot spot single base mutations. The potentially mutated bases are underlined (A, G, C and T) in the synthetic template. The synthetic template is immobilized on a

References

Antao VP, Lai SY, Tinoco I Jr. (1991) A thermodynamic study of unusually stable RNA and DNA hairpins. *Nucleic Acids Res.* **19**: 5901-5905.

Axelrod VD, Vartikyan RM, Aivazashvili VA, Beabealashvili RS. (1978) Specific termination of RNA polymerase synthesis as a method of RNA and DNA sequencing. *Nucleic Acids Res.* **5**(10): 3549-3563.

Badman ER and Cooks RG. (2000) Cylindrical Ion Trap Array with Mass Selection by Variation in Trap Dimensions *Anal. Chem.* 72(20):5079-5086.

Badman ER and Cooks RG. (2000) A Parallel Miniature Cylindrical Ion Trap Array. *Anal. Chem.* 72(14):3291-3297.

Bowling JM, Bruner KL, Cmarik JL, Tibbetts C. (1991) Neighboring nucleotide interactions during DNA sequencing gel electrophoresis. *Nucleic Acids Res.* **19**: 3089-3097.

Burgess K, Jacutin SE, Lim D, Shitangkoon A. (1997) An approach to photolabile, fluorescent protecting groups. *J. Org. Chem.* **62(15)**: 5165-5168.

Canard B, Cardona B, Sarfati RS. (1995) Catalytic editing properties of DNA polymerases. *Proc. Natl. Acad. Sci. USA* **92**: 10859-10863.

chromophores in energy transfer primers. *Anal Biochem.* **243**(1): 15-27.

Hyman ED, (1988) A new method of sequencing DNA. *Analytical Biochemistry* **174**: 423-436.

Ireland RE, Varney MD (1986) Approach to the total synthesis of chlorothricolide-synthesis of (+/-)-19.20-dihydro-24-O-methylchlorothricolide, methyl-ester, ethyl carbonate. *J. Org. Chem.* **51**: 635-648.

Ju J, Glazer AN, Mathies RA. (1996) Cassette labeling for facile construction of energy transfer fluorescent primers. *Nucleic Acids Res.* **24**: 1144-1148.

Ju J, Ruan C, Fuller CW, Glazer AN Mathies RA. (1995) Energy transfer fluorescent dye-labeled primers for DNA sequencing and analysis. *Proc. Natl. Acad. Sci. USA* **92**: 4347-4351.

Kamal A, Laxman E, Rao NV. (1999) A mild and rapid regeneration of alcohols from their allylic ethers by chlorotrimethylsilane/sodium iodide. *Tetrahedron letters* **40**: 371-372.

Kheterpal I, Scherer J, Clark SM, Radhakrishnan A, Ju J, Ginther CL, Sensabaugh GF, Mathies RA. (1996) DNA Sequencing Using a Four-Color Confocal Fluorescence Capillary Array Scanner. *Electrophoresis.* **17**: 1852-1859.

Characterization of the p53 tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. *Cancer Res.* **57**: 4285-4300.

Olejnik J, Ludemann HC, Krzymanska-Olejnik E, Berkenkamp S, Hillenkamp F, Rothschild KJ. (1999) Photocleavable peptide-DNA conjugates: synthesis and applications to DNA analysis using MALDI-MS. *Nucleic Acids Res.* **27**: 4626-4631.

Olejnik J, Sonar S, Krzymanska-Olejnik E, Rothschild KJ. (1995) Photocleavable biotin derivatives: a versatile approach for the isolation of biomolecules. *Proc. Natl. Acad. Sci. USA.* **92**: 7590-7594.

Pelletier H, Sawaya MR, Kumar A, Wilson SH, Kraut J. (1994) Structures of ternary complexes of rat DNA polymerase ß, a DNA template-primer, and ddCTP. *Science* **264**: 1891-1903.

Pennisi E. (2000) DOE Team Sequences Three Chromosomes. *Science* **288**: 417 - 419.

Pillai VNR. (1980) Photoremovable Protecting Groups in Organic Synthesis. *Synthesis* 1-62.

Prober JM, Trainor GL, Dam RJ, Hobbs FW, Robertson CW, Zagursky RJ, Cocuzza AJ, Jensen MA, Baumeister K. (1987) A system for rapid DNA sequencing with fluorescent

Simpson PC, Adam DR, Woolley T, Thorsen T, Johnston R, Sensabaugh GF, and Mathies RA. (1998) High-throughput genetic analysis using microfabricated 96-sample capillary array electrophoresis microplates. *Proc. Natl. Acad. Sci. U. S. A.* 95:2256-2261.

Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SBH, Hood LE. (1986) Fluorescence detection in automated DNA sequencing analysis. *Nature* **321**: 674-679.

Tabor S, Richardson C.C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **84**: 4767-4771.

Tabor S. & Richardson, CC. (1995) A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **92**: 6339-6343.

Turro NJ. (1991) Modern Molecular Photochemistry; University Science Books, Mill Valley, CA.

Velculescu VE, Zhang, I, Vogelstein, B. and Kinzler KW (1995) Serial Analysis of Gene Expression. *Science* **270**: 484-487.

Welch MB, Burgess K, (1999) Synthesis of fluorescent, photolabile 3'-O-protected nucleoside triphosphates for the base addition sequencing scheme. *Nucleosides and Nucleotides* **18**:197-201.

What is claimed is:

1. A method for sequencing a nucleic acid by detecting the identity of a nucleotide analogue after the nucleotide analogue is incorporated into a growing strand of DNA in a polymerase reaction, which comprises the following steps:

(i) attaching a 5′ end of the nucleic acid to a solid surface;

(ii) attaching a primer to the nucleic acid attached to the solid surface;

(iii) adding a polymerase and one or more different nucleotide analogues to the nucleic acid to thereby incorporate a nucleotide analogue into the growing strand of DNA, wherein the incorporated nucleotide analogue terminates the polymerase reaction and wherein each different nucleotide analogue comprises (a) a base selected from the group consisting of adenine, guanine, cytosine, thymine, and uracil, and their analogues; (b) a unique label attached through a cleavable linker to the base or to an analogue of the base; (c) a deoxyribose; and (d) a cleavable chemical group to cap an -OH group at a 3′-position of the deoxyribose;

(iv) washing the solid surface to remove unincorporated nucleotide analogues;

wherein if the unique label is a dye, the order of steps (v) through (vii) is: (v), (vi), and (vii); and

wherein if the unique label is a mass tag, the order of steps (v) through (vii) is: (vi), (vii), and (v).

2. The method of claim 1, wherein the solid surface is glass, silicon, or gold.

3. The method of claim 1, wherein the solid surface is a magnetic bead, a chip, a channel in a chip, or a porous channel in a chip.

4. The method of claim 1, wherein the step of attaching the nucleic acid to the solid surface comprises:

(i) coating the solid surface with a phosphine moiety,

(ii) attaching an azido group to the 5′ end of the nucleic acid, and

(iii) immobilizing the 5′ end of the nucleic acid to the solid surface through interaction between the phosphine moiety on the solid surface and the azido group on the 5′ end of the nucleic acid.

deoxyribose capable of self-priming in the polymerase reaction.

10. The method of claim 1, wherein the step of attaching the primer to the nucleic acid comprises hybridizing the primer to the nucleic acid or ligating the primer to the nucleic acid.

11. The method of claim 1, wherein one or more of four different nucleotide analogues is added in step (iii), wherein each different nucleotide analogue comprises a different base selected from the group consisting of thymine or uracil or an analogue of thymine or uracil, adenine or an analogue of adenine, cytosine or an analogue of cytosine, and guanine or an analogue of guanine, and wherein each of the four different nucleotide analogues comprises a unique label.

12. The method of claim 1, wherein the cleavable chemical group that caps the -OH group at the 3'-position of the deoxyribose in the nucleotide analogue is $-CH_2OCH_3$ or $-CH_2CH=CH_2$.

13. The method of claim 1, wherein the unique label that is attached to the nucleotide analogue is a fluorescent moiety or a fluorescent semiconductor crystal.

14. The method of claim 13, wherein the fluorescent moiety is selected from the group consisting of 5-carboxyfluorescein, 6-carboxyrhodamine-6G,

of cytosine or thymine or to a 7-position of deaza-adenine or deaza-guanine.

20. The method of claim 1, wherein the cleavable linker between the unique label and the nucleotide analogue is cleaved by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

21. The method of claim 20, wherein the cleavable linker is a photocleavable linker which comprises a 2-nitrobenzyl moiety.

22. The method of claim 1, wherein the cleavable chemical group used to cap the -OH group at the 3'-position of the deoxyribose is cleaved by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

23. The method of claim 1, wherein the chemical compounds added in step (vi) to permanently cap any unreacted -OH group on the primer attached to the nucleic acid or on the primer extension strand are a polymerase and one or more different dideoxynucleotides or analogues of dideoxynucleotides.

24. The method of claim 23, wherein the different dideoxynucleotides are selected from the group consisting of 2',3'-dideoxyadenosine 5'-

simultaneously applying the method of claim 1 to the plurality of different nucleic acids.

28. Use of the method of claim 1 or 27 for detection of single nucleotide polymorphisms, genetic mutation analysis, serial analysis of gene expression, gene expression analysis, identification in forensics, genetic disease association studies, DNA sequencing, genomic sequencing, translational analysis, or transcriptional analysis.

29. A method of attaching a nucleic acid to a solid surface which comprises:

(i) coating the solid surface with a phosphine moiety,

(ii) attaching an azido group to a 5′ end of the nucleic acid, and

(iii) immobilizing the 5′ end of the nucleic acid to the solid surface through interaction between the phosphine moiety on the solid surface and the azido group on the 5′ end of the nucleic acid.

30. The method of claim 29, wherein the step of coating the solid surface with the phosphine moiety comprises:

(i) coating the surface with a primary amine, and

(a) a base selected from the group consisting of adenine or an analogue of adenine, cytosine or an analogue of cytosine, guanine or an analogue of guanine, thymine or an analogue of thymine, and uracil or an analogue of uracil;

(b) a unique label attached through a cleavable linker to the base or to an analogue of the base;

(c) a deoxyribose; and

(d) a cleavable chemical group to cap an -OH group at a 3'-position of the deoxyribose.

38. The nucleotide analogue of claim 37, wherein the cleavable chemical group that caps the -OH group at the 3'-position of the deoxyribose is $-CH_2OCH_3$ or $-CH_2CH=CH_2$.

39. The nucleotide analogue of claim 37, wherein the unique label is a fluorescent moiety or a fluorescent semiconductor crystal.

40. The nucleotide analogue of claim 39, wherein the fluorescent moiety is selected from the group consisting of 5-carboxyfluorescein, 6-carboxyrhodamine-6G, N,N,N',N'-tetramethyl-6-carboxyrhodamine, and 6-carboxy-X-rhodamine.

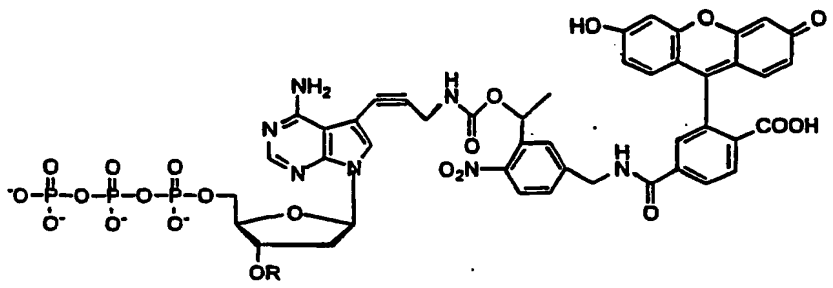
41. The nucleotide analogue of claim 37, wherein the unique label is a fluorescence energy transfer tag

means, a chemical means, a physical chemical means, heat, and light.

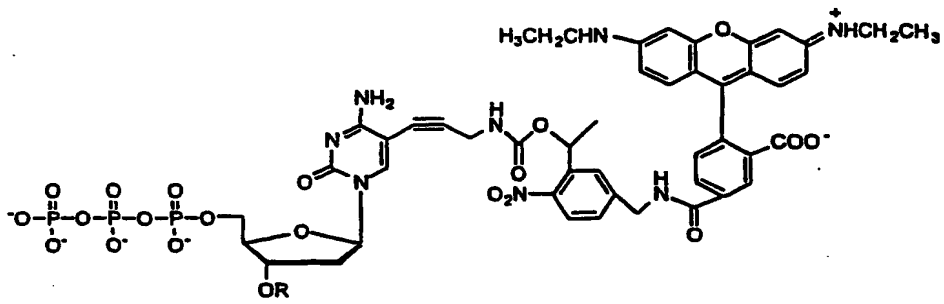
47. The nucleotide analogue of claim 46, wherein the cleavable linker is a photocleavable linker which comprises a 2-nitrobenzyl moiety.

48. The nucleotide analogue of claim 37, wherein the cleavable chemical group used to cap the -OH group at the 3'-position of the deoxyribose is cleavable by a means selected from the group consisting of one or more of a physical means, a chemical means, a physical chemical means, heat, and light.

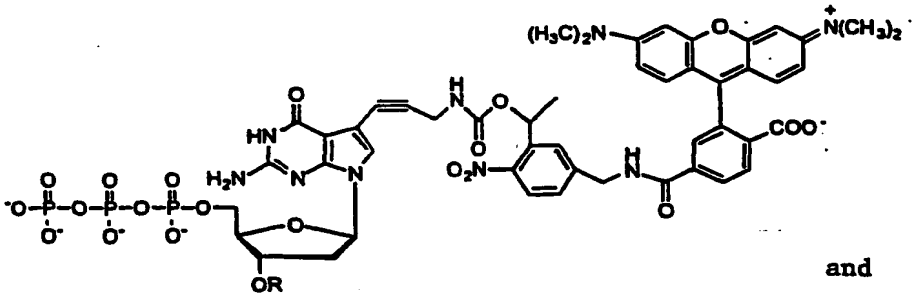
50. The nucleotide analogue of claim 49, wherein the nucleotide analogue is selected from the group consisting of:



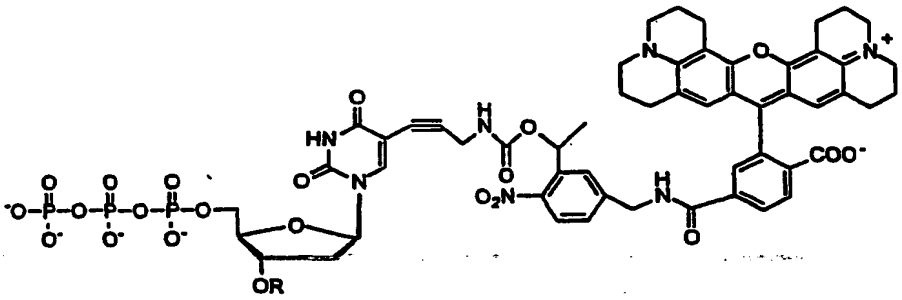
,



,



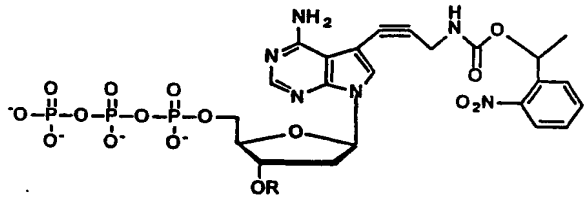
and

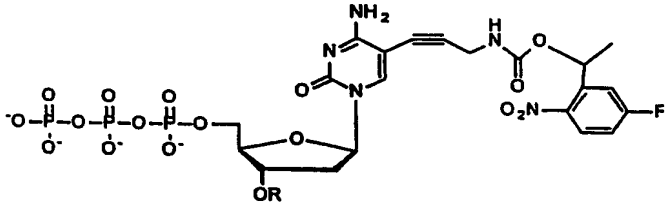


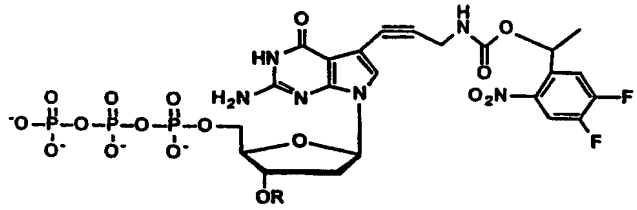
;

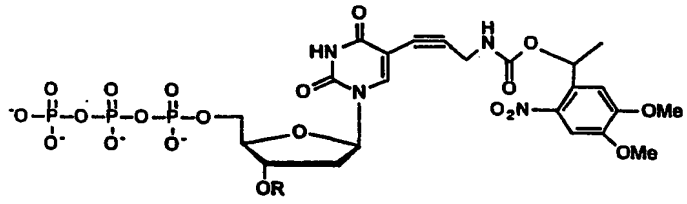
wherein R is $-CH_2OCH_3$ or $-CH_2CH=CH_2$.

52. The nucleotide analogue of claim 51, wherein the nucleotide analogue is selected from the group consisting of:


,


,


, and

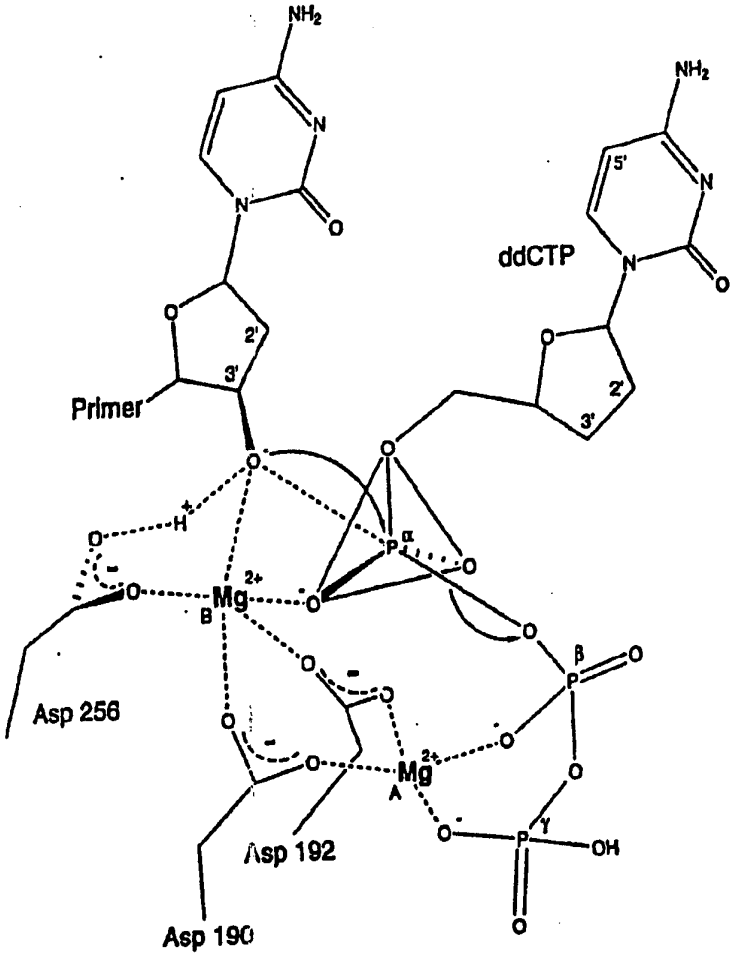

;

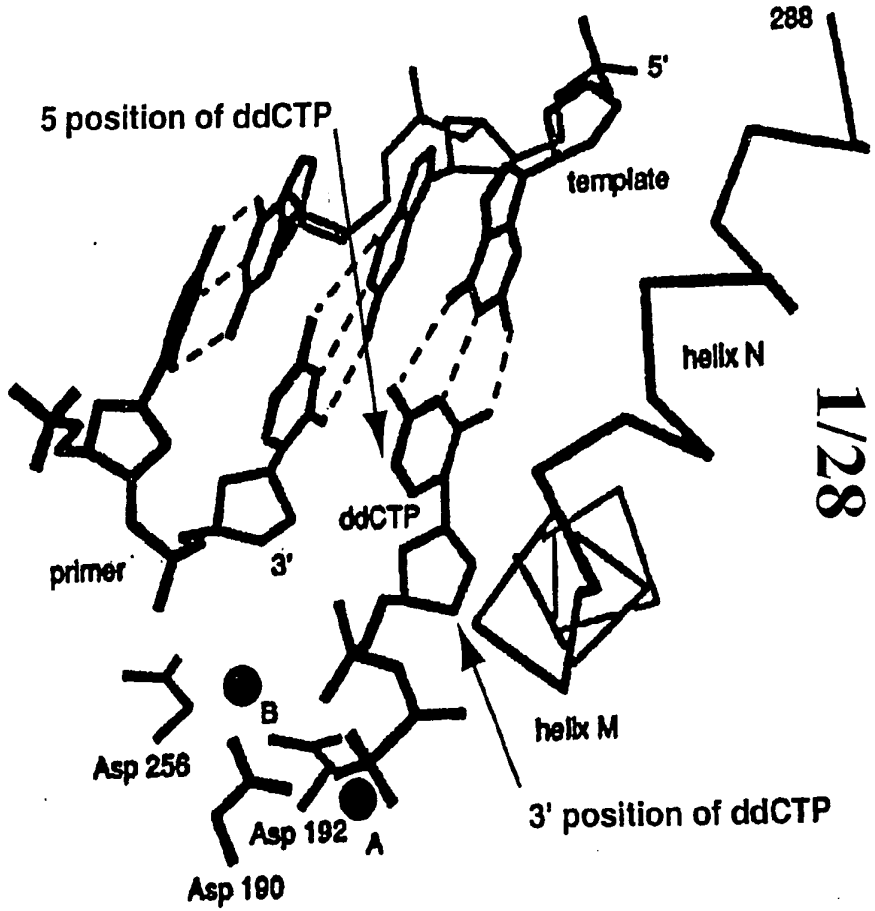
wherein R is $-CH_2OCH_3$ or $-CH_2CH=CH_2$.

59. The system of claim 54, wherein the mass tags have molecular weights between 150 daltons and 250 daltons.

60. Use of the system of claim 54 for DNA sequencing analysis, detection of single nucleotide polymorphisms, genetic mutation analysis, serial analysis of gene expression, gene expression analysis, identification in forensics, genetic disease association studies, DNA sequencing, genomic sequencing, translational analysis, or transcriptional analysis.

FIGURE 1





Asp = Aspartic Acid

FIGURE 2A

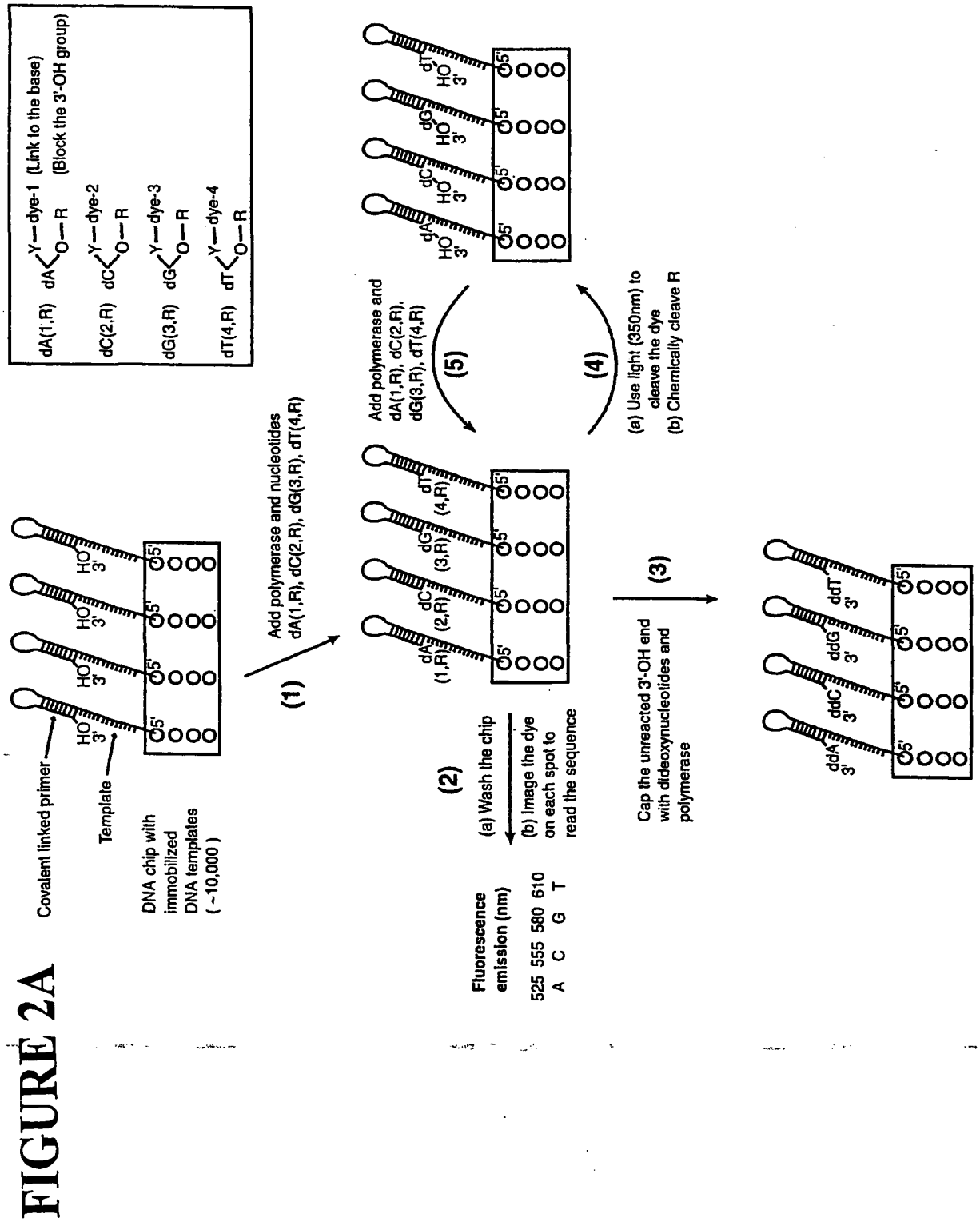

dA(1,R) dA Y—dye-1 O—R (Link to the base) (Block the 3'-OH group)
dC(2,R) dC Y—dye-2 O—R
dG(3,R) dG Y—dye-3 O—R
dT(4,R) dT Y—dye-4 O—R
Covalent linked primer
Template
DNA chip with immobilized DNA templates (~10,000)
(1)
Add polymerase and nucleotides dA(1,R), dC(2,R), dG(3,R), dT(4,R)
(2)
(a) Wash the chip
(b) Image the dye on each spot to read the sequence
Fluorescence emission (nm)
525 555 580 610
A C G T
(3)
Cap the unreacted 3'-OH end with dideoxynucleolides and polymerase
(4)
(a) Use light (350nm) to cleave the dye
(b) Chemically cleave R
(5)
Add polymerase and dA(1,R), dC(2,R), dG(3,R), dT(4,R)

FIGURE 2B

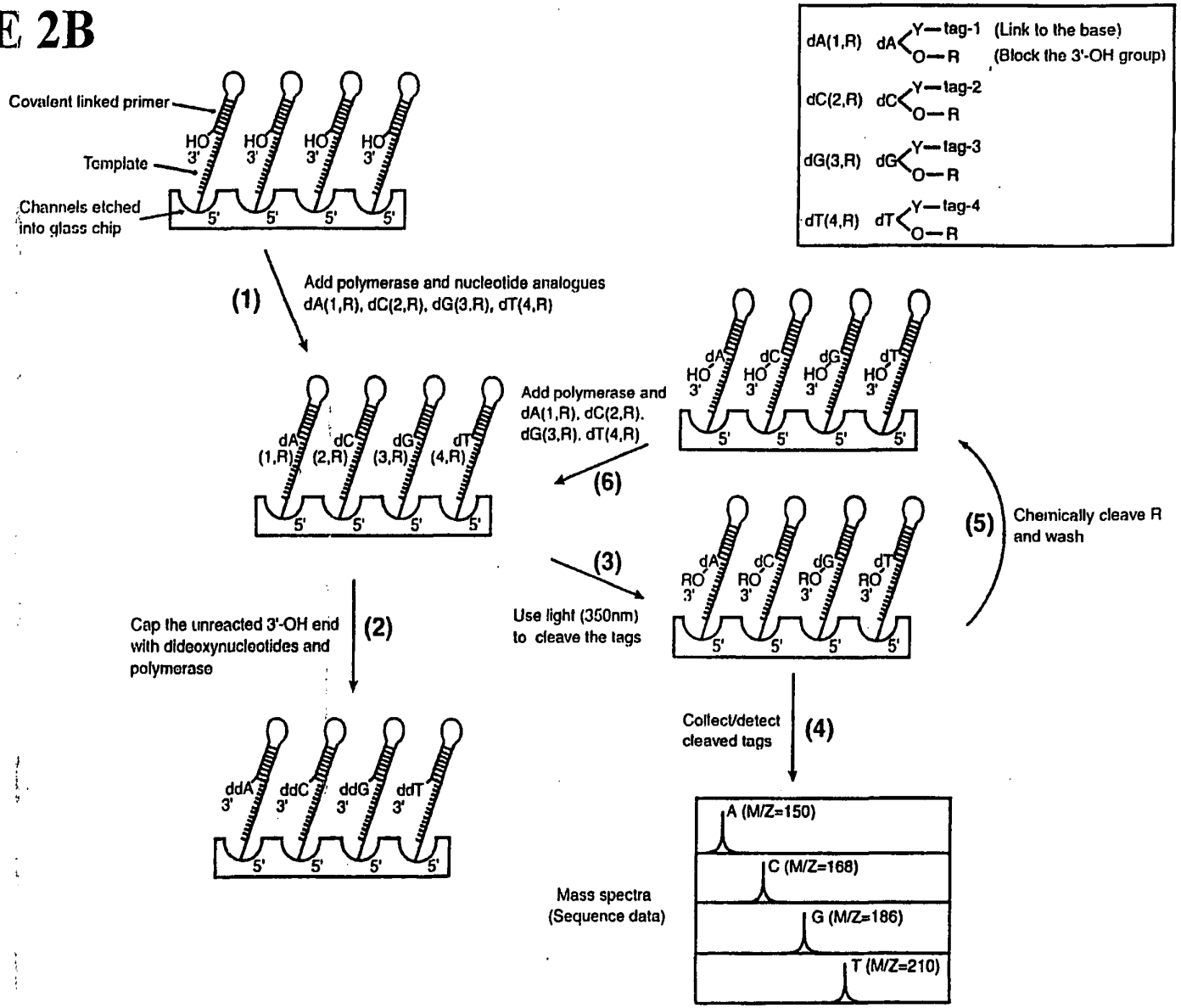

dA(1,R) dA Y—tag-1 O—R (Link to the base) (Block the 3'-OH group)
dC(2,R) dC Y—tag-2 O—R
dG(3,R) dG Y—tag-3 O—R
dT(4,R) dT Y—tag-4 O—R
Covalent linked primer
Template
Channels etched into glass chip
HO 3'
5'
(1)
Add polymerase and nucleotide analogues dA(1,R), dC(2,R), dG(3,R), dT(4,R)
dA (1,R)
dC (2,R)
dG (3,R)
dT (4,R)
Add polymerase and dA(1,R), dC(2,R), dG(3,R), dT(4,R)
(6)
(3)
Use light (350nm) to cleave the tags
RO 3'
(5)
Chemically cleave R and wash
Cap the unreacted 3'-OH end with dideoxynucleotides and polymerase
(2)
ddA 3'
ddC 3'
ddG 3'
ddT 3'
Collect/detect cleaved tags
(4)
A (M/Z=150)
C (M/Z=168)
G (M/Z=186)
T (M/Z=210)
Mass spectra (Sequence data)

FIGURE 3

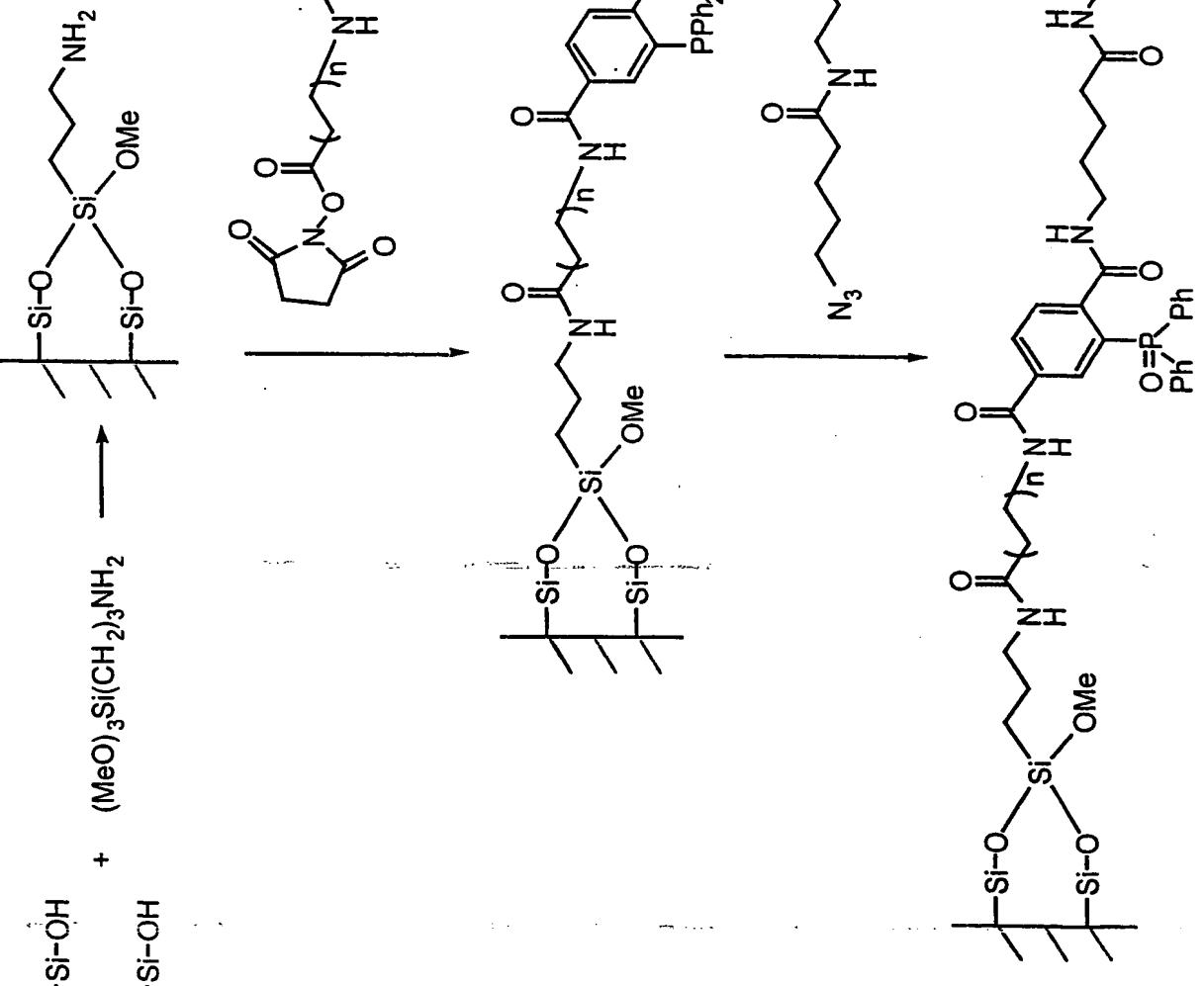
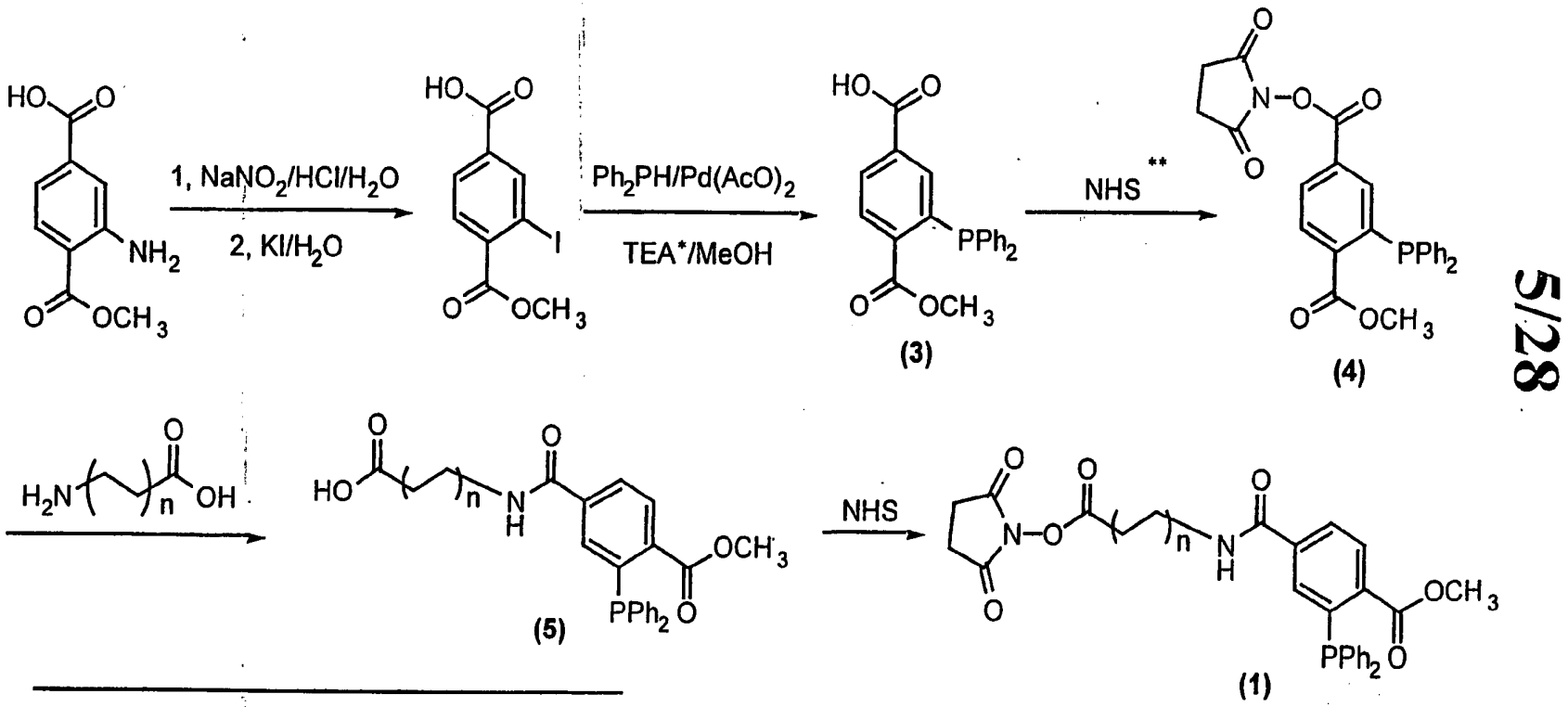
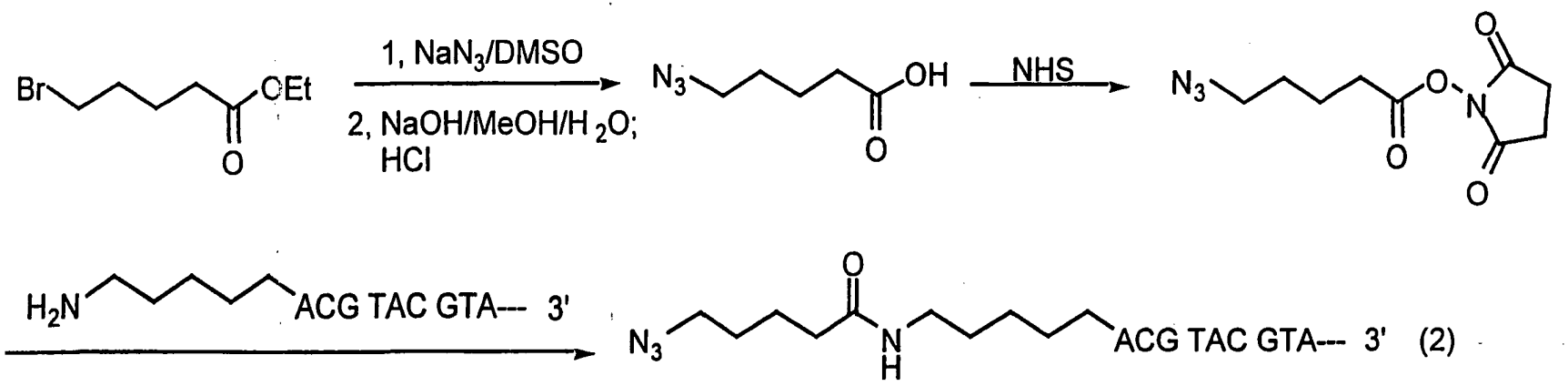

Si-OH
Si-OH
+
(MeO)3Si(CH2)3NH2
NH2
OMe
Si
O
(1)
OCH3
PPh2
n
ACG TAC GTA--- 3'
(2)
N3
R
Ph

FIGURE 4



*TEA = Triethylamine;**NHS = N-Hydroxysuccinimide

FIGURE 5


Br
CEt
O
1, NaN3/DMSO
2, NaOH/MeOH/H 2O;
HCl
N3
OH
O
NHS
N3
O
O
N
O
O
H2N
ACG TAC GTA--- 3'
N3
O
N
H
ACG TAC GTA--- 3'
(2)

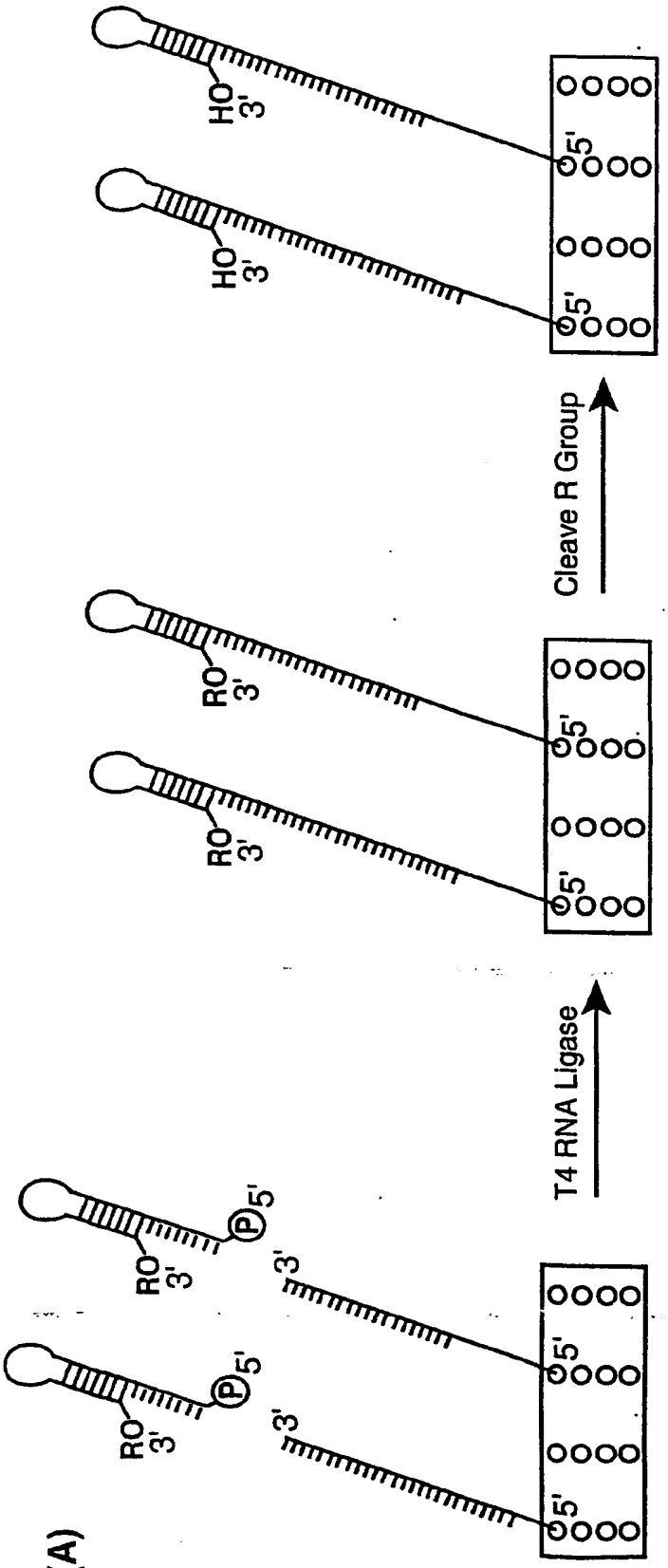


FIGURE 6A

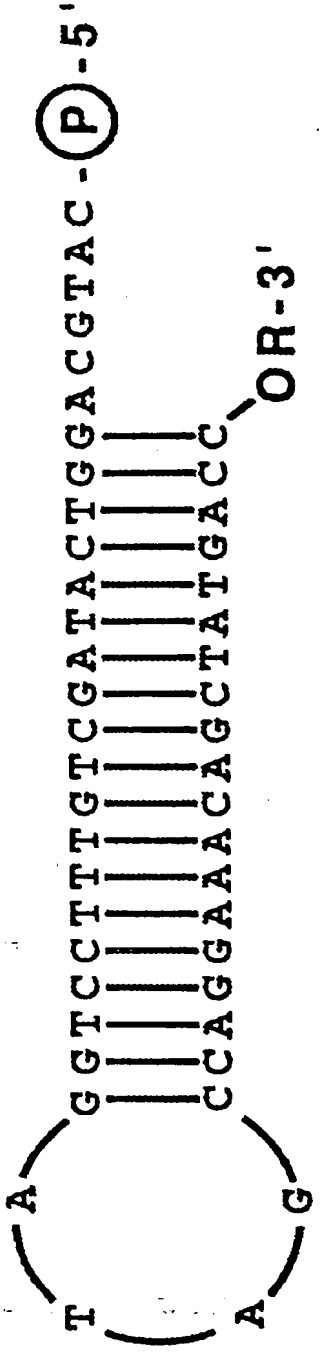
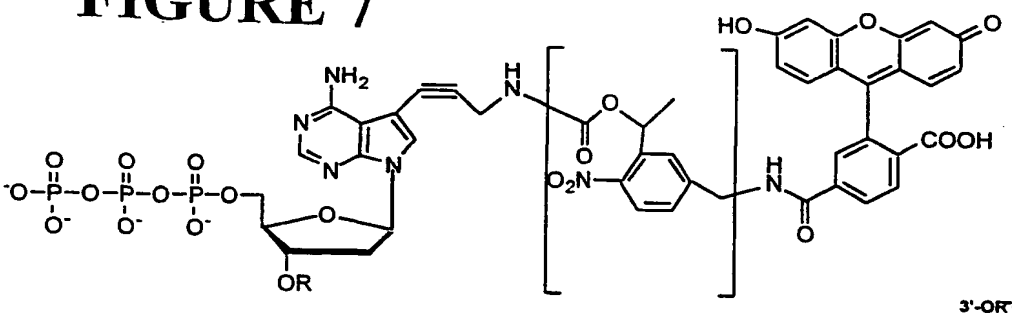

(B)
5'-(P)-
GGTCCTTTGTCGATACTGGACGTAC-
CCAGGAAACAGCTATGACC-
OR-3'
A
G
T
A

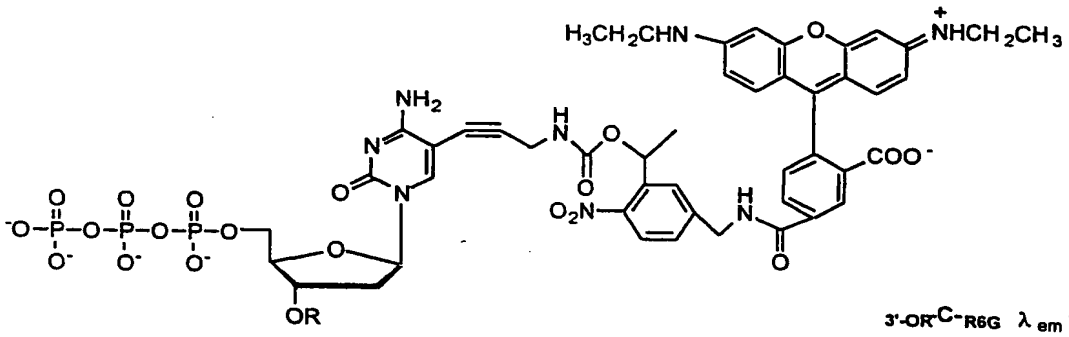
FIGURE 6B

FIGURE 7

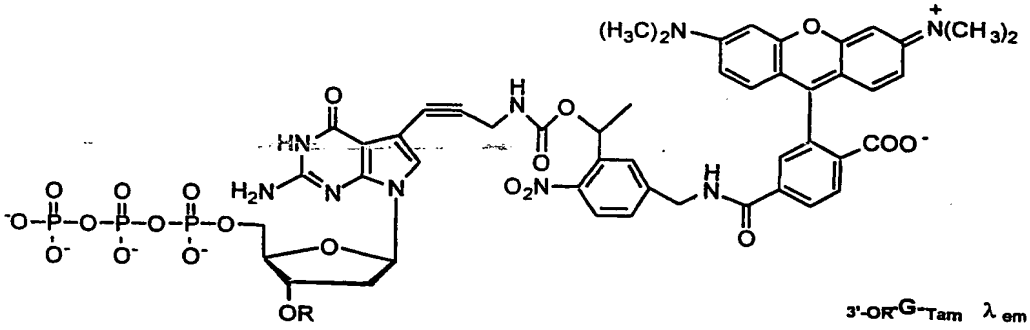


3'-OR^{A}-$_{Fam}$ λ_{em} = 525 nm

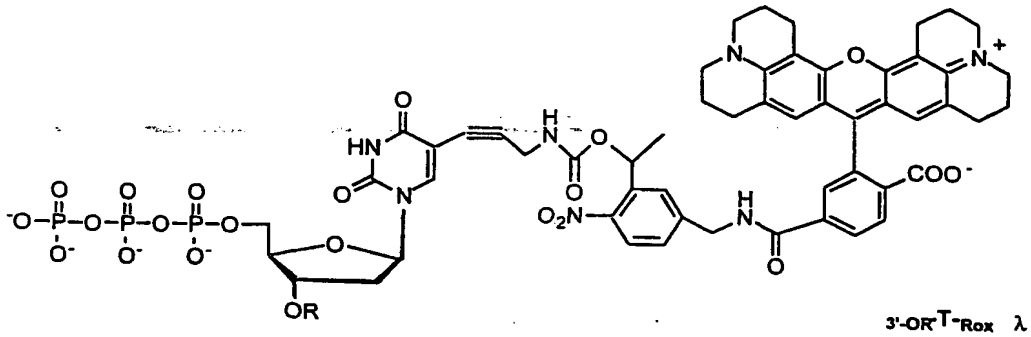
Photocleavable Linker



3'-OR^{C}-$_{R6G}$ λ_{em} = 555 nm



3'-OR^{G}-$_{Tam}$ λ_{em} = 585 nm



3'-OR^{T}-$_{Rox}$ λ_{em} = 604 nm

R = H, CH_2OCH_3 (MOM) or CH_2-CH=CH_2 (Allyl)

FIGURE 8

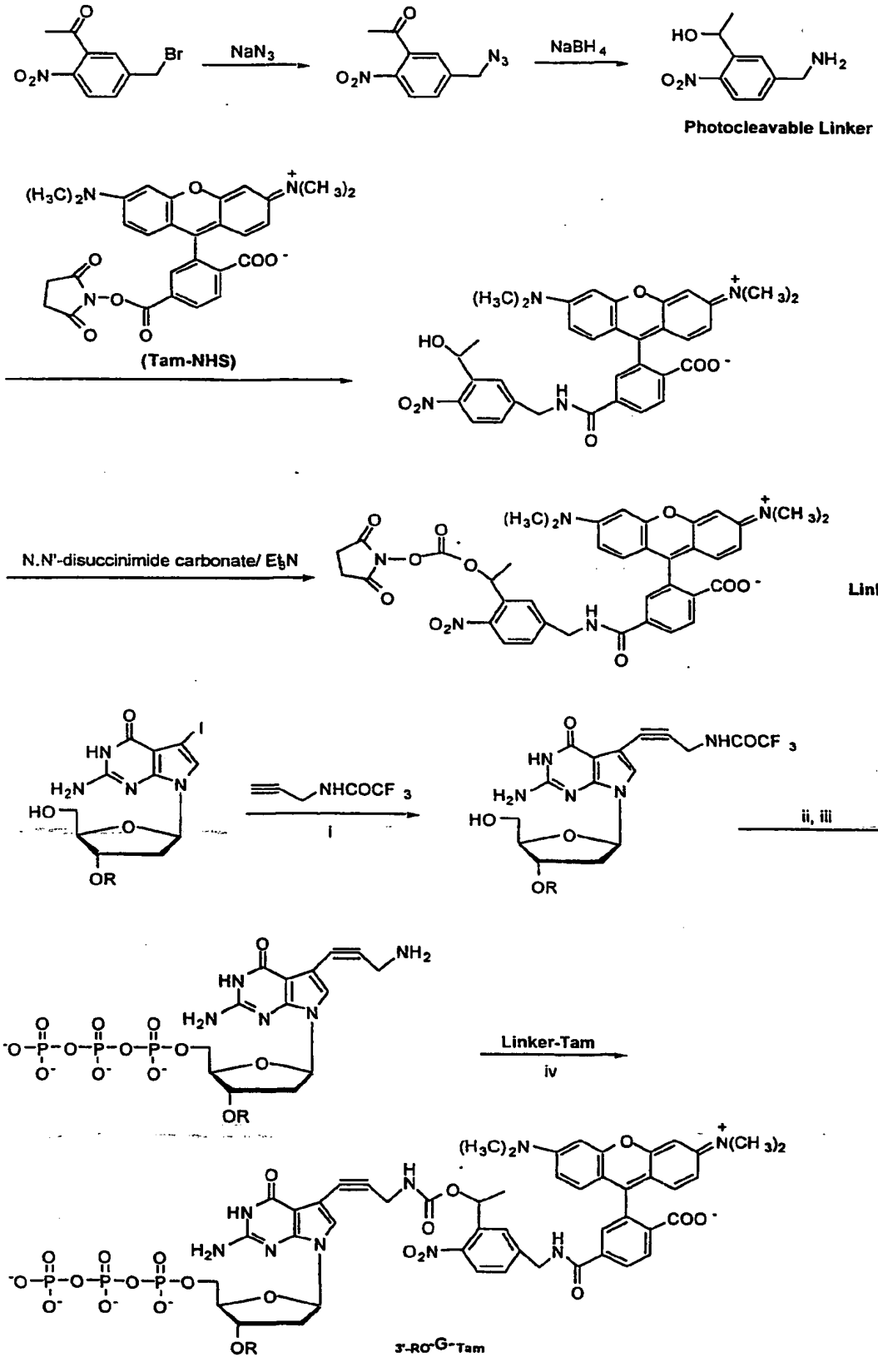

NaN3
NaBH4
Photocleavable Linker
(Tam-NHS)
N.N'-disuccinimide carbonate/ Et3N
Linker-Tam
NHCOCF 3
i
ii, iii
iv
Linker-Tam
3'-RO-G-Tam

FIGURE 9

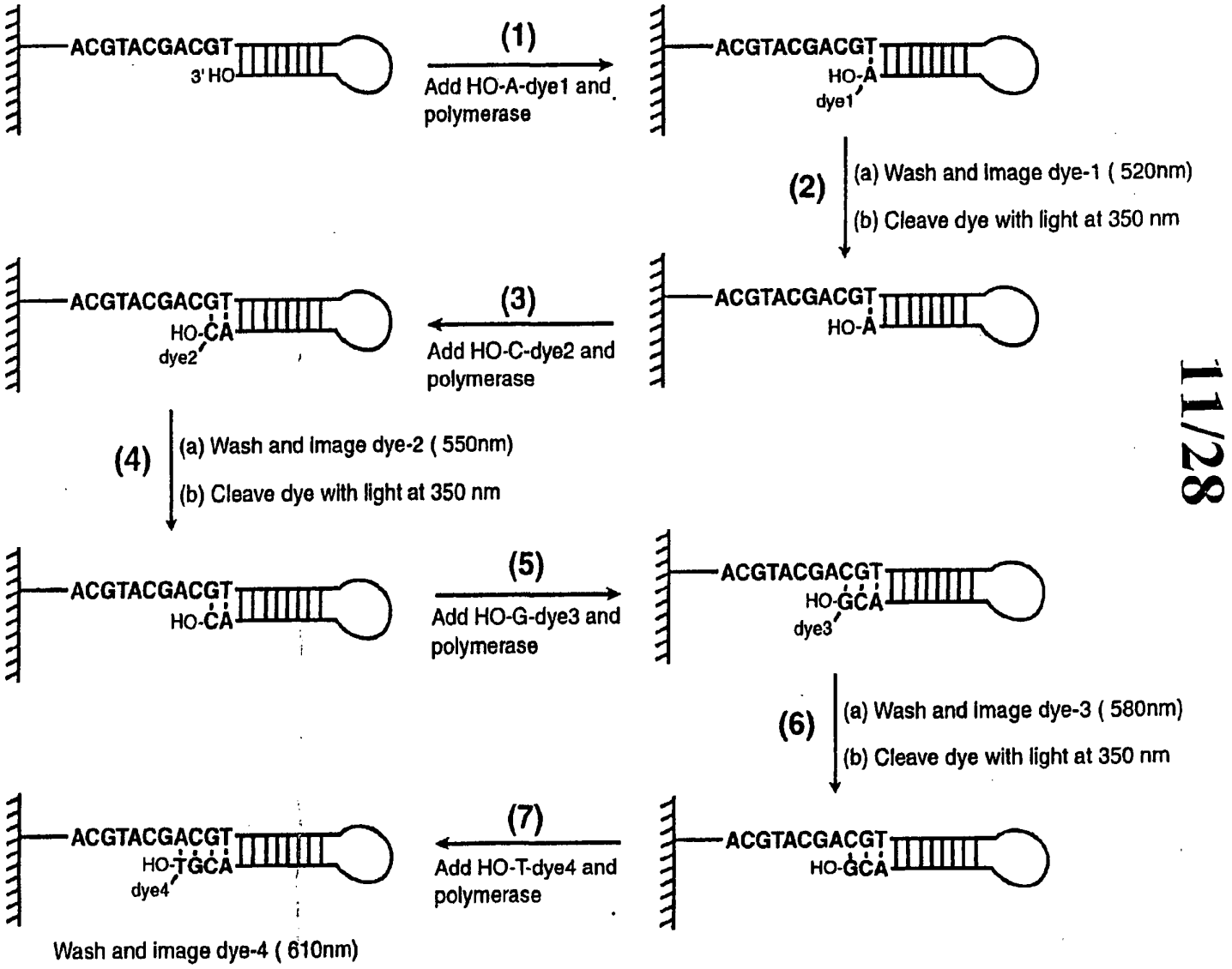
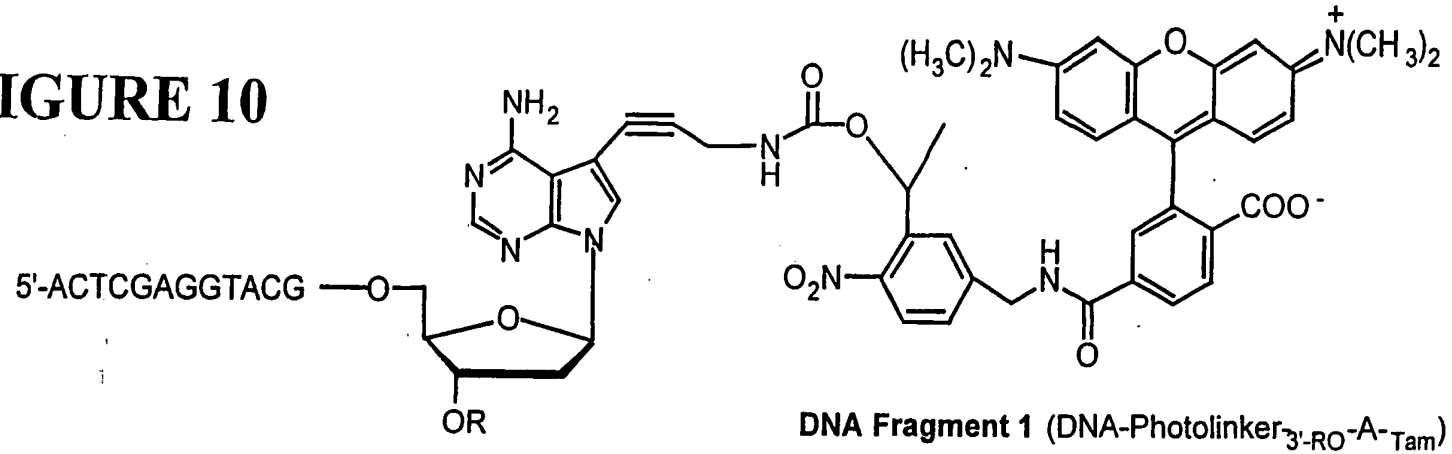
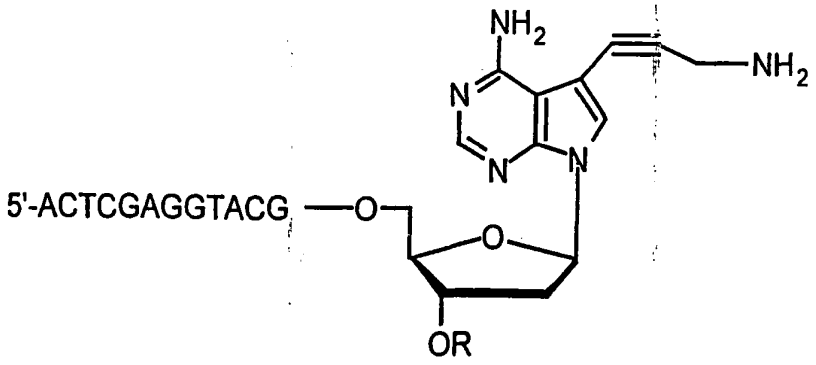

ACGTACGACGT
3' HO
(1)
Add HO-A-dye1 and polymerase
HO-A
dye1
(2)
(a) Wash and Image dye-1 (520nm)
(b) Cleave dye with light at 350 nm
HO-A
(3)
Add HO-C-dye2 and polymerase
HO-CA
dye2
(4)
(a) Wash and Image dye-2 (550nm)
(b) Cleave dye with light at 350 nm
HO-CA
(5)
Add HO-G-dye3 and polymerase
HO-GCA
dye3
(6)
(a) Wash and Image dye-3 (580nm)
(b) Cleave dye with light at 350 nm
HO-GCA
(7)
Add HO-T-dye4 and polymerase
HO-TGCA
dye4
Wash and image dye-4 (610nm)

FIGURE 10



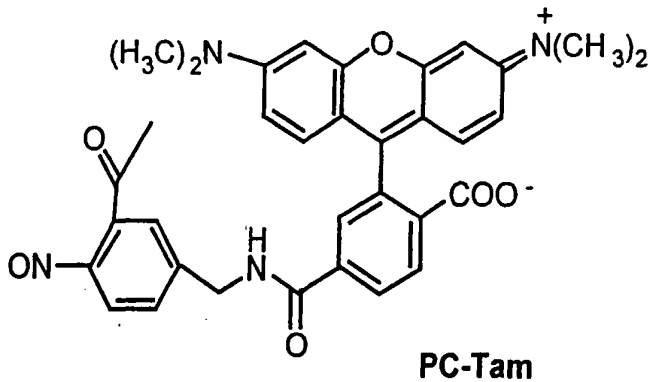
DNA Fragment 1 (DNA-Photolinker$_{3'\text{-RO}}$-A-$_{Tam}$)

$h\nu$ (300-360nm)



DNA Fragment 2

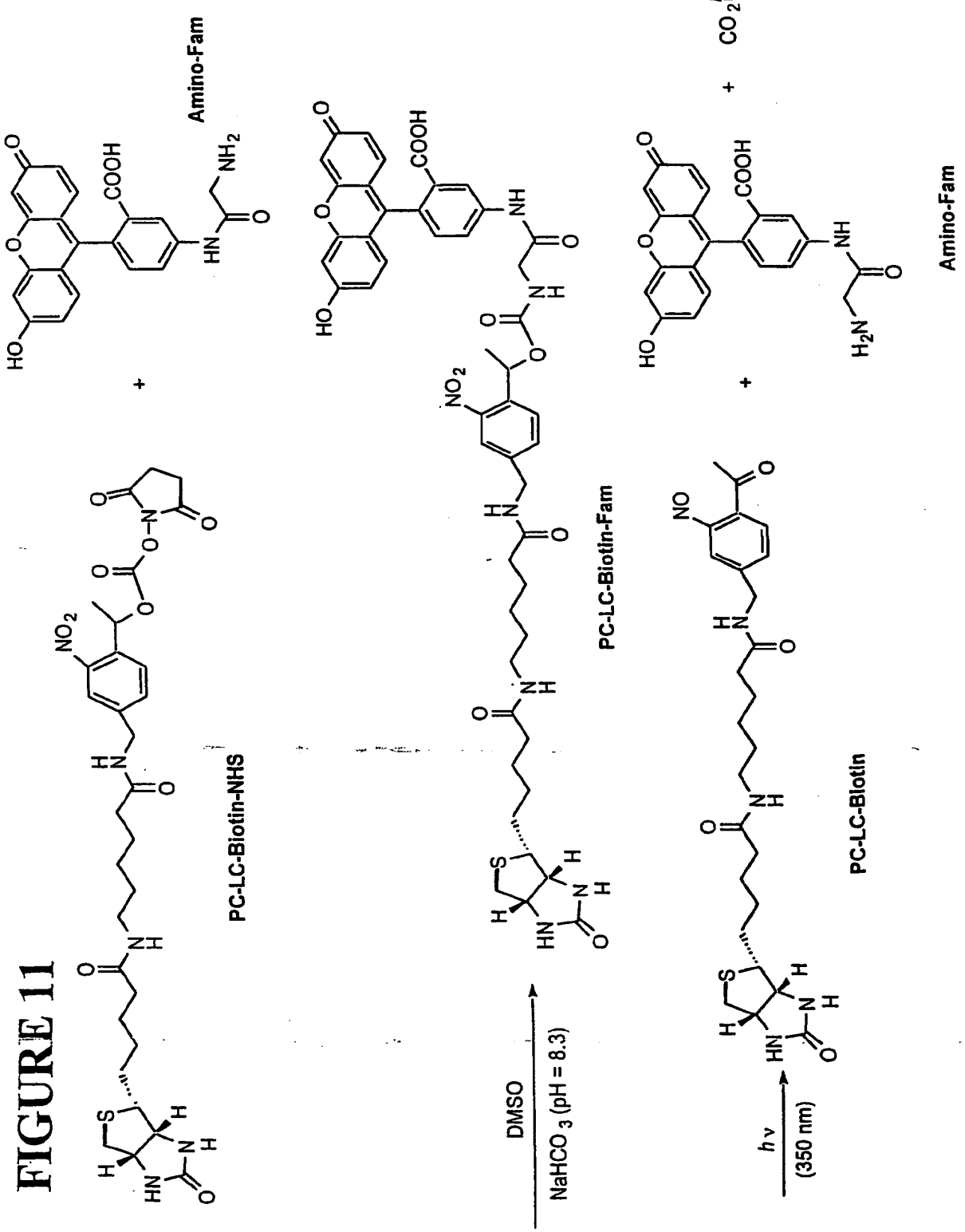
\+



PC-Tam

\+ CO_2

FIGURE 11


Amino-Fam
PC-LC-Biotin-NHS
DMSO
NaHCO3 (pH = 8.3)
PC-LC-Biotin-Fam
hν
(350 nm)
PC-LC-Biotin
CO2
Amino-Fam

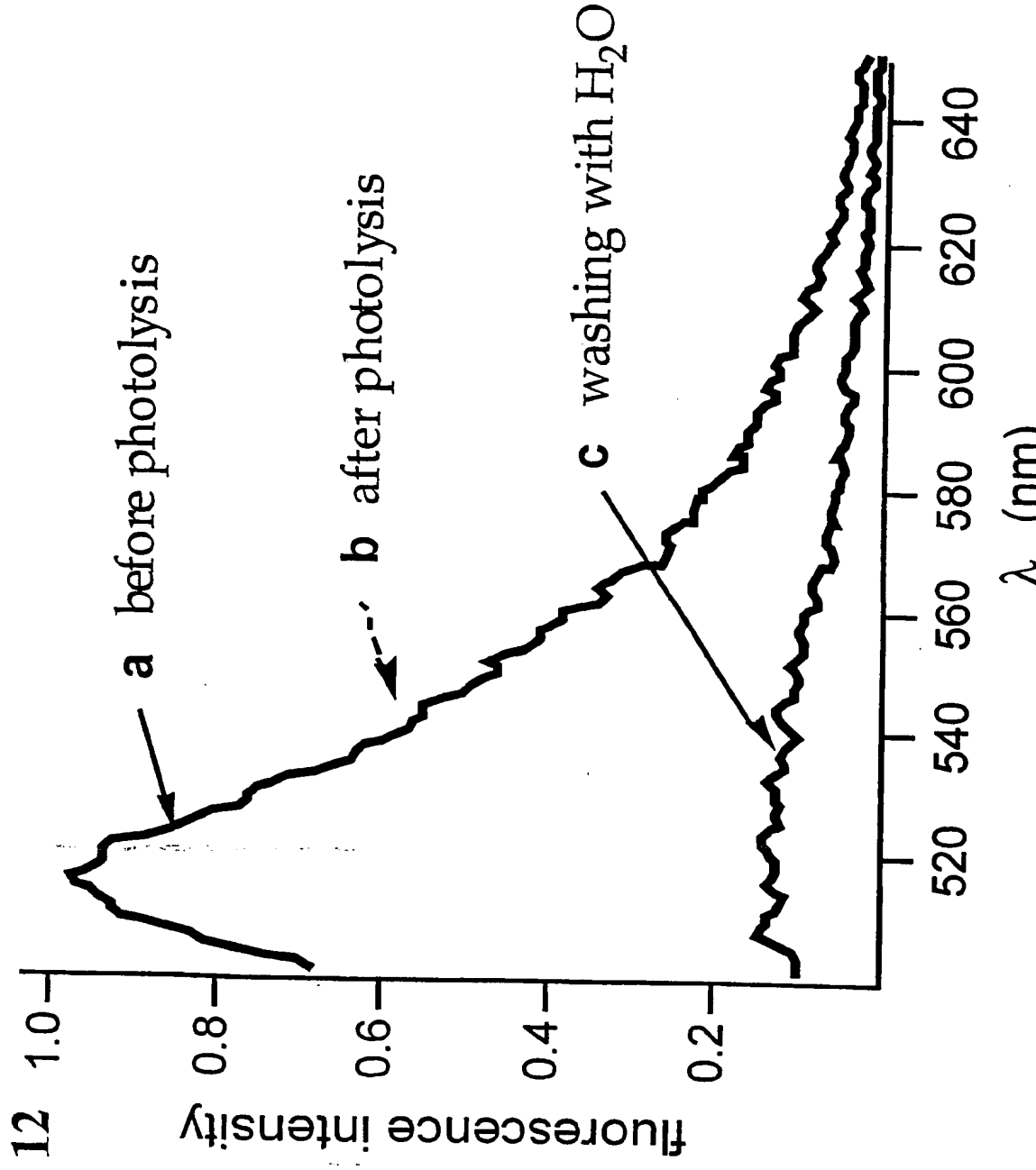
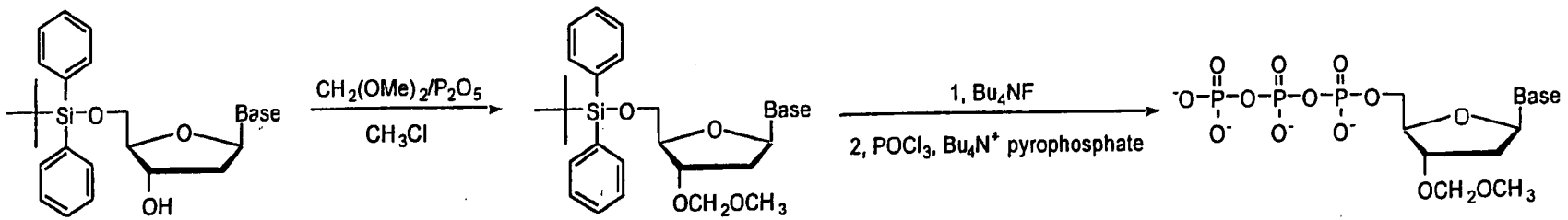

a before photolysis
b after photolysis
c washing with H_2O
fluorescence intensity
1.0
0.8
0.6
0.4
0.2
520
540
560
580
600
620
640
λ (nm)

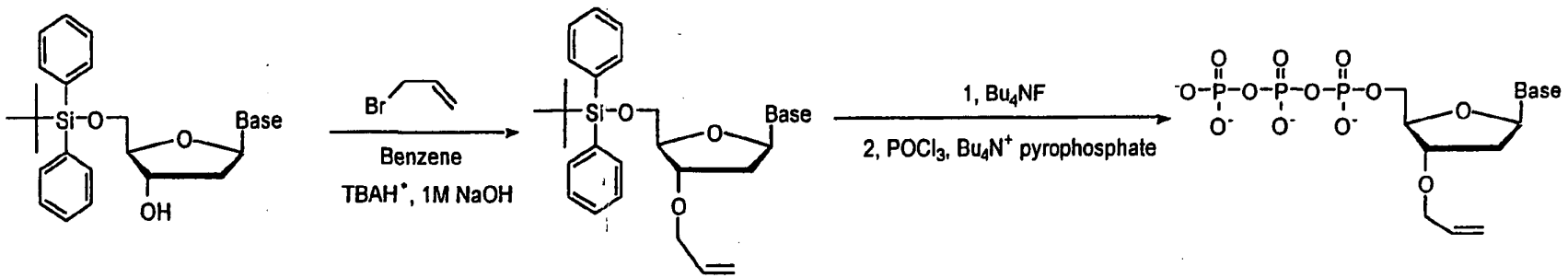
FIGURE 12

FIGURE 13A


Si
O
Base
OH
$CH_2(OMe)_2/P_2O_5$
CH_3Cl
OCH_2OCH_3
1, Bu_4NF
2, $POCl_3$, Bu_4N^+ pyrophosphate

(A)

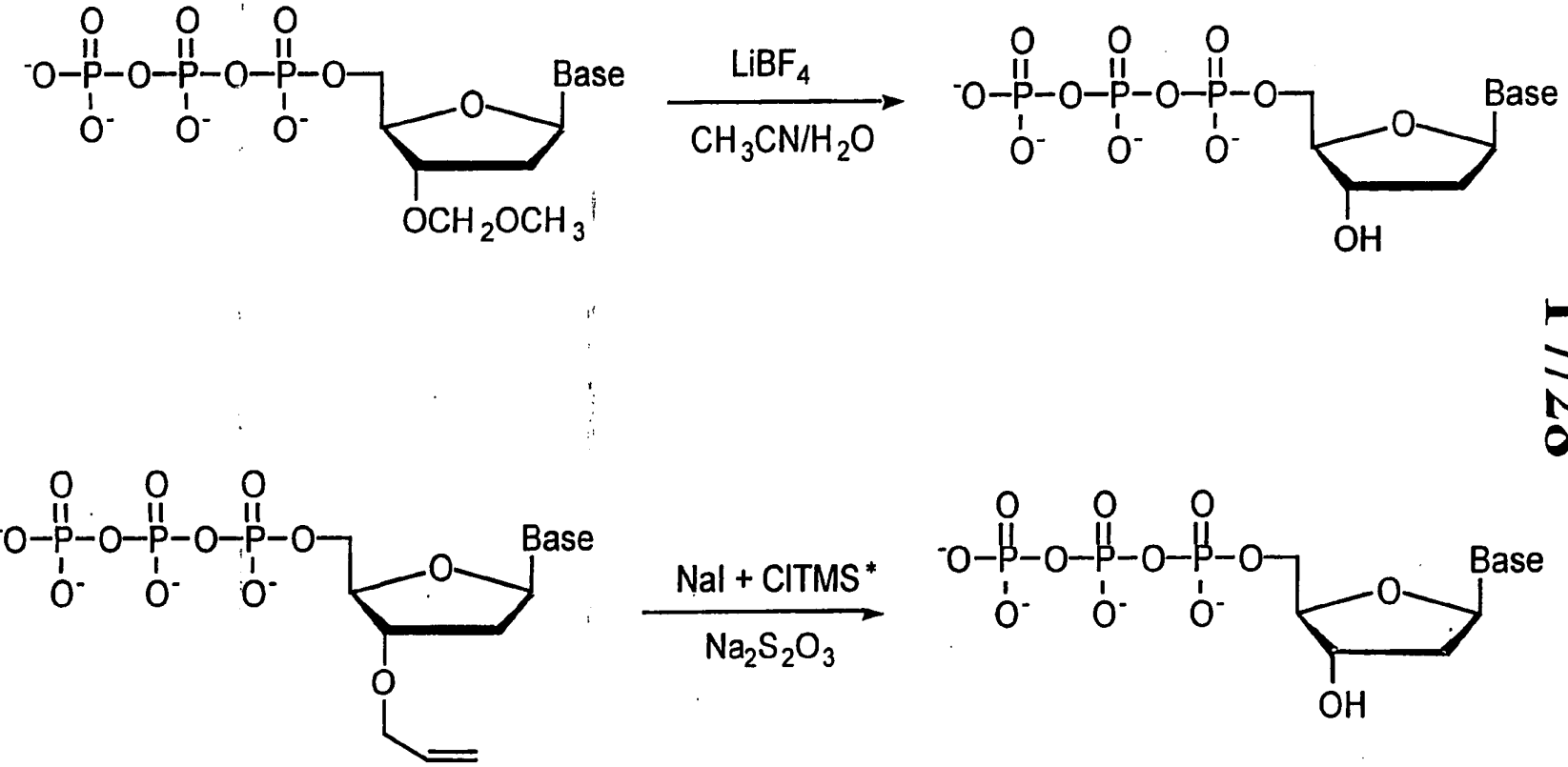
FIGURE 13B



(B)

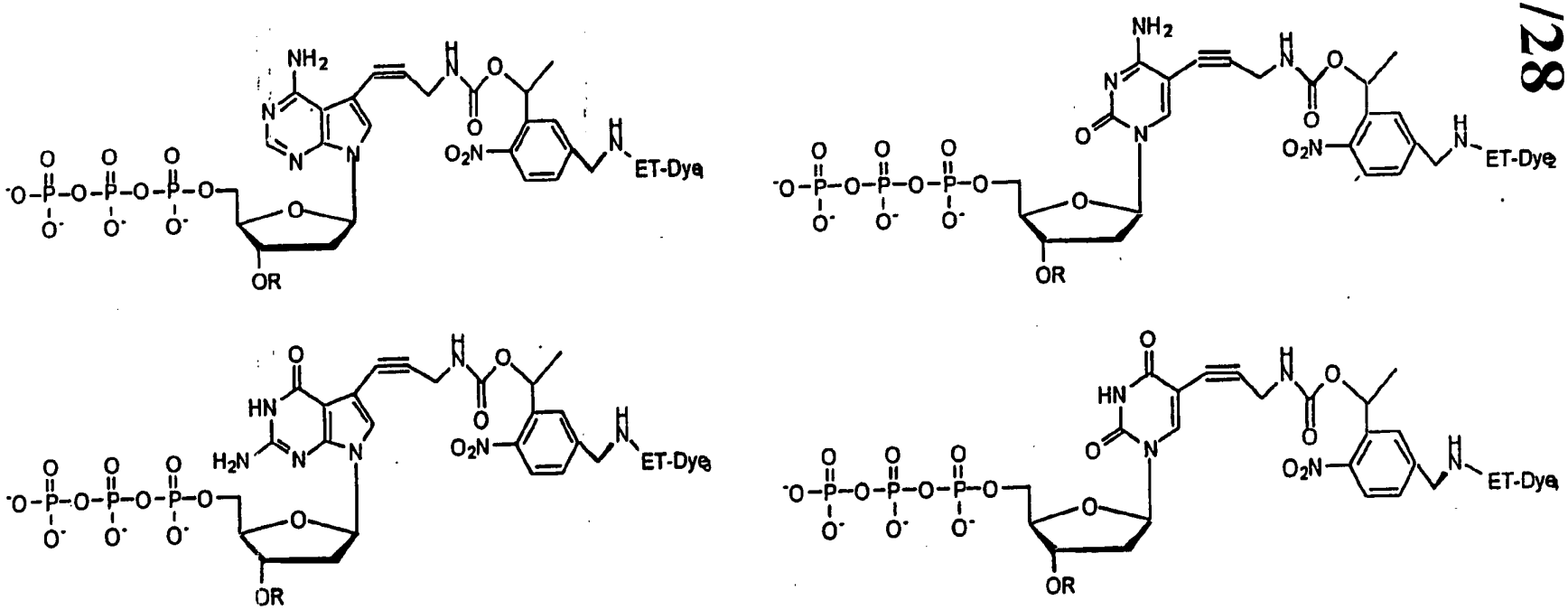
* TBAH = Tetrabutylammonium hydroxide

FIGURE 14



*CITMS = chlorotrimethylsilane

FIGURE 15A



R = H , CH_2OCH_3 (MOM) or CH_2-CH=CH_2 (Allyl)

FIGURE 15B

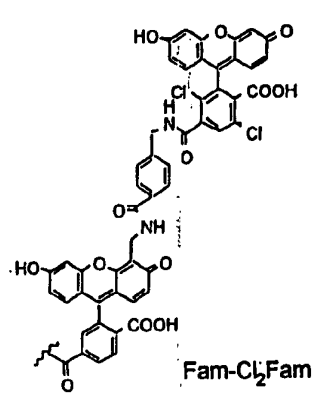
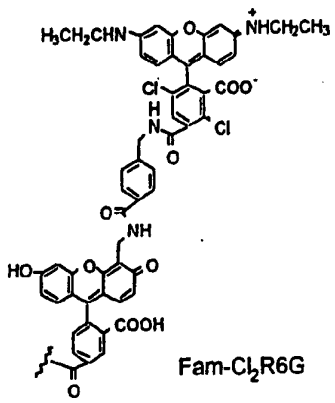
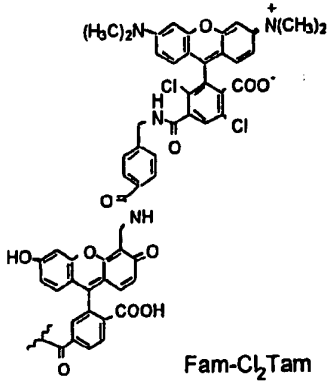
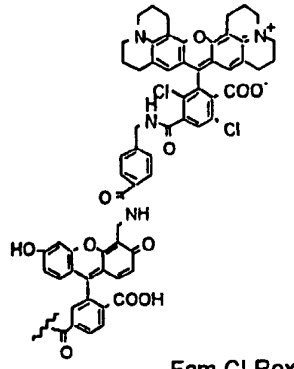
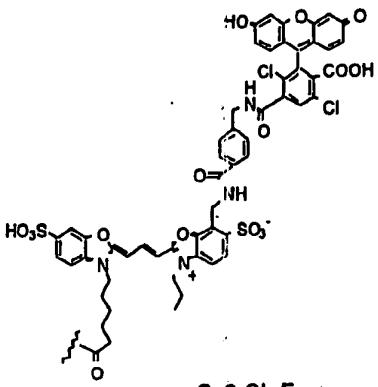
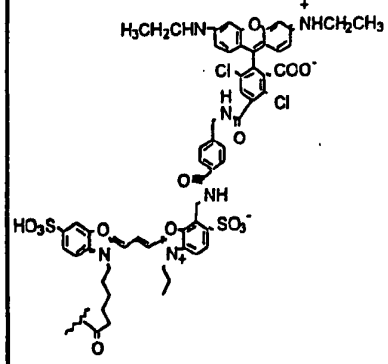
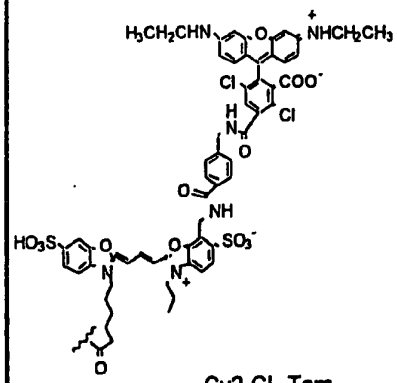
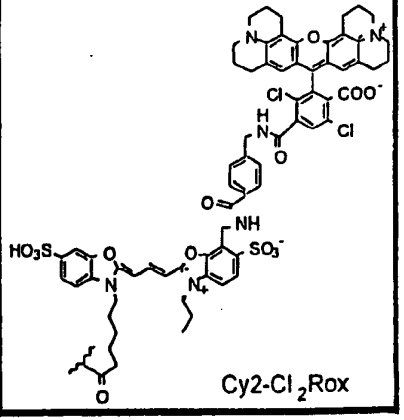
ET Dye$_1$	ET Dye$_2$	ET Dye$_3$	ET Dye$_4$
λ_{em} = 530 nm	λ_{em} = 560 nm	λ_{em} = 590 nm	λ_{em} = 620 nm
 Fam-Cl$_2$Fam	 Fam-Cl$_2$R6G	 Fam-Cl$_2$Tam	 Fam-Cl$_2$Rox
 Cy2-Cl$_2$Fam	 Cy2-Cl$_2$R6G	 Cy2-Cl$_2$Tam	 Cy2-Cl$_2$Rox

FIGURE 16

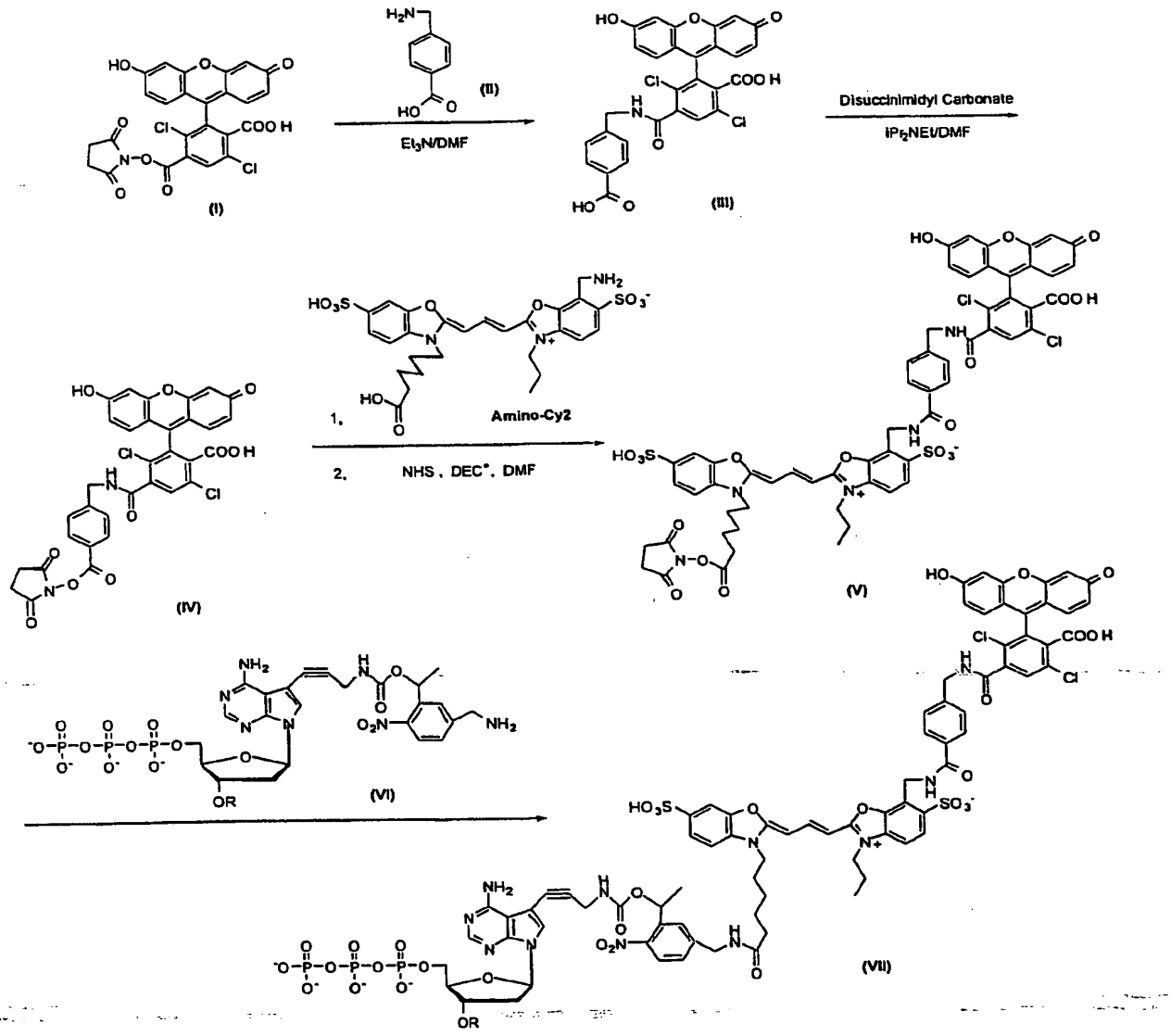

(I)
(II)
Et3N/DMF
(III)
Disuccinimidyl Carbonate
iPr2NEt/DMF
Amino-Cy2
1.
2. NHS, DEC*, DMF
(IV)
(V)
(VI)
OR
(VII)

FIGURE 17

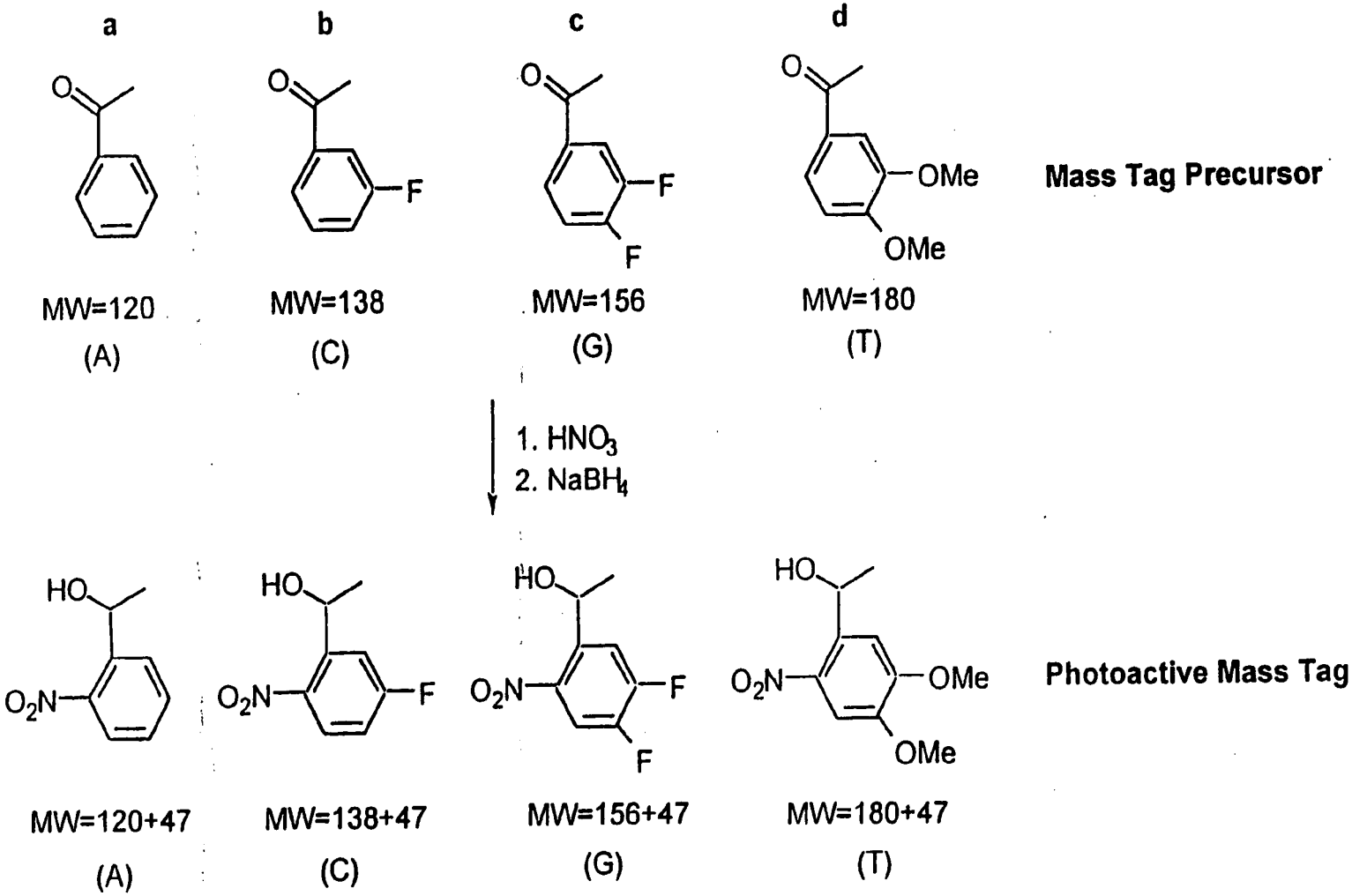

a
b
c
d
MW=120
(A)
MW=138
(C)
MW=156
(G)
MW=180
(T)
Mass Tag Precursor
1. HNO3
2. NaBH4
MW=120+47
(A)
MW=138+47
(C)
MW=156+47
(G)
MW=180+47
(T)
Photoactive Mass Tag

FIGURE 18

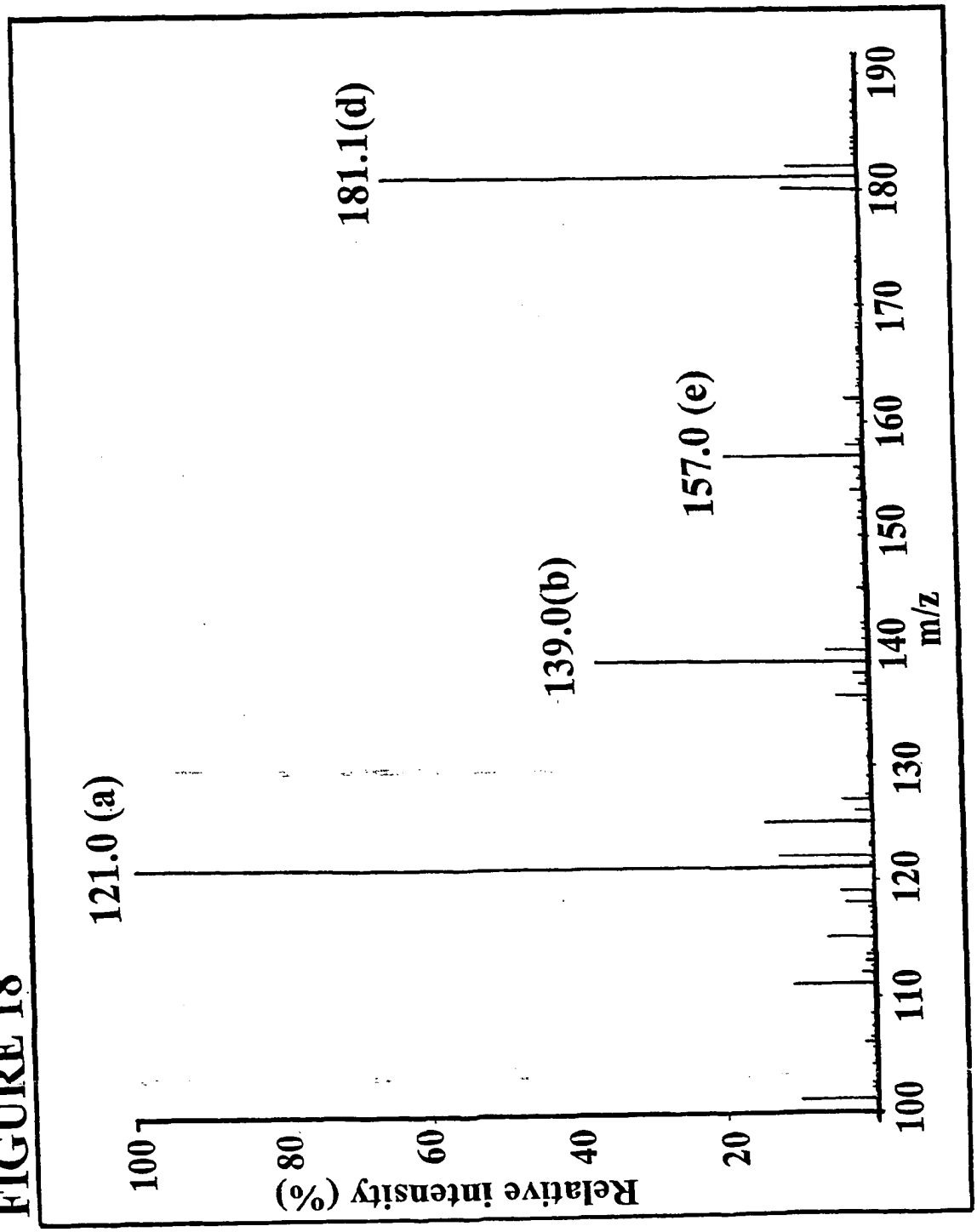
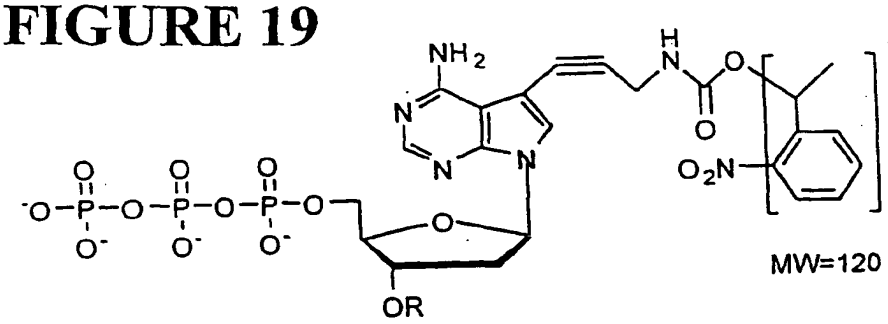
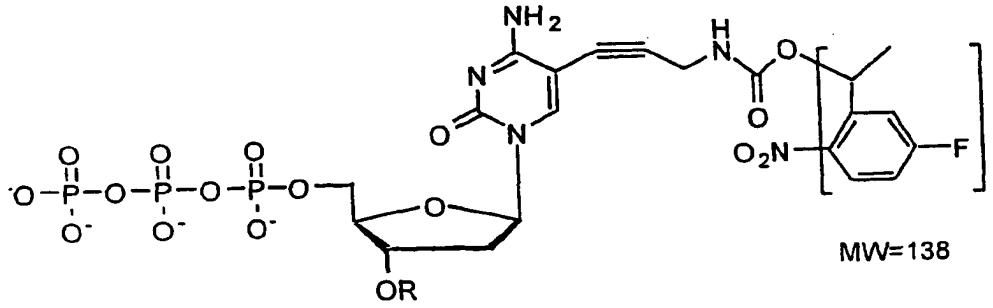

Relative intensity (%)
100
80
60
40
20
100
110
120
130
140
150
160
170
180
190
m/z
121.0 (a)
139.0(b)
157.0 (e)
181.1(d)

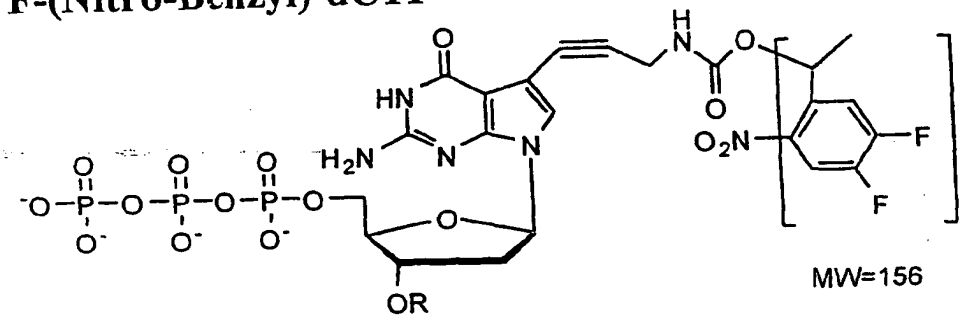
FIGURE 19



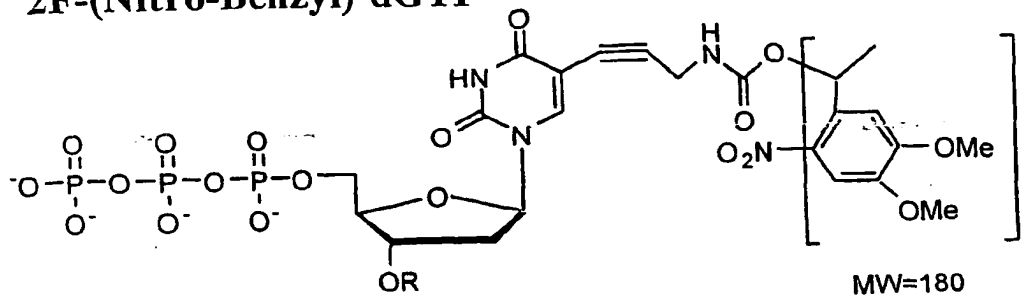
(Nitro-Benzyl)-dATP



F-(Nitro-Benzyl)-dCTP



2F-(Nitro-Benzyl)-dGTP



2(Meo)-(Nitro-Benzyl)-dTTP

R = H, MOM or Allyl

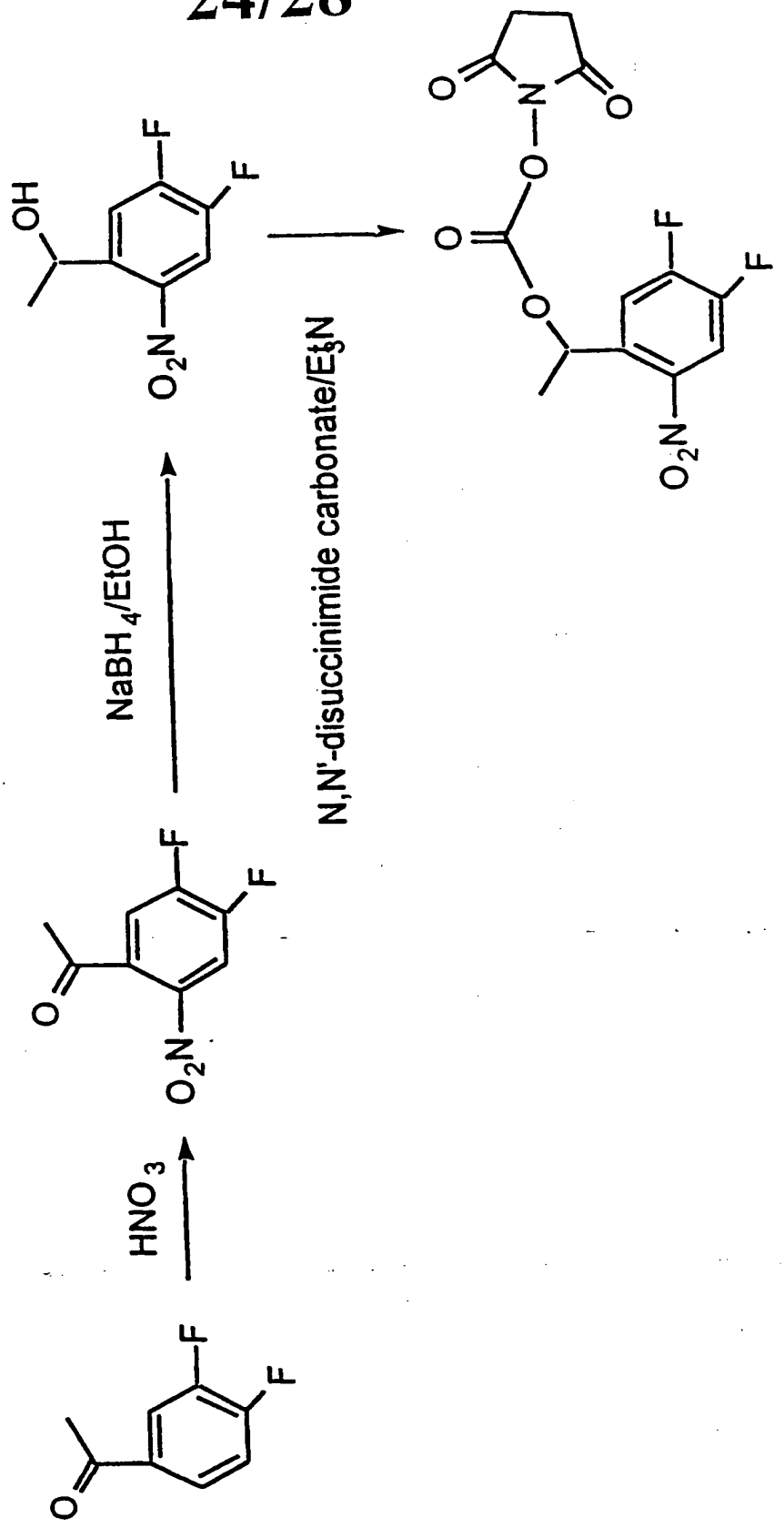

HNO3
NaBH4/EtOH
N,N'-disuccinimide carbonate/Et3N
F
F
O
O2N
OH
O
N

FIGURE 20

FIGURE 21

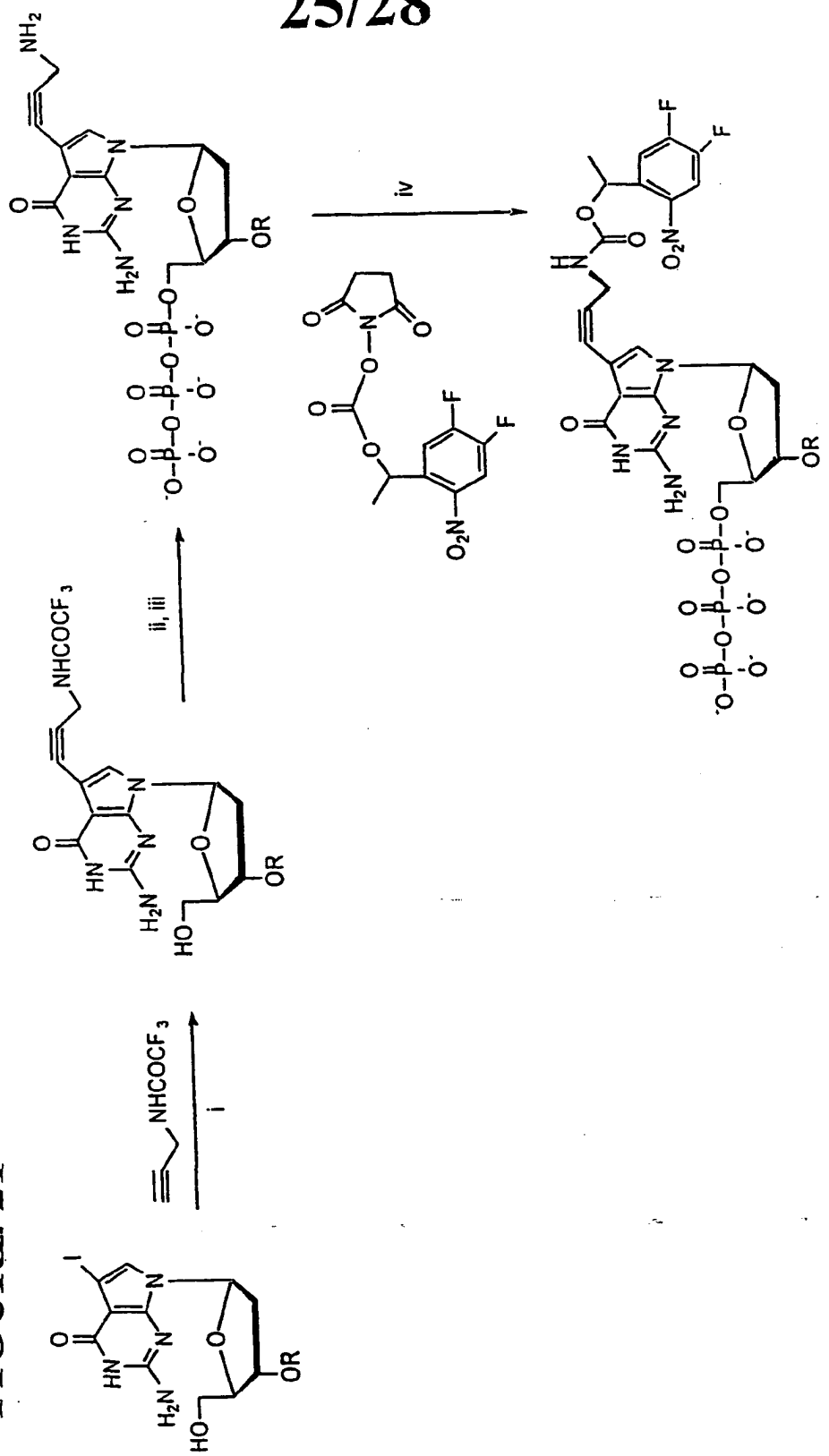


FIGURE 22

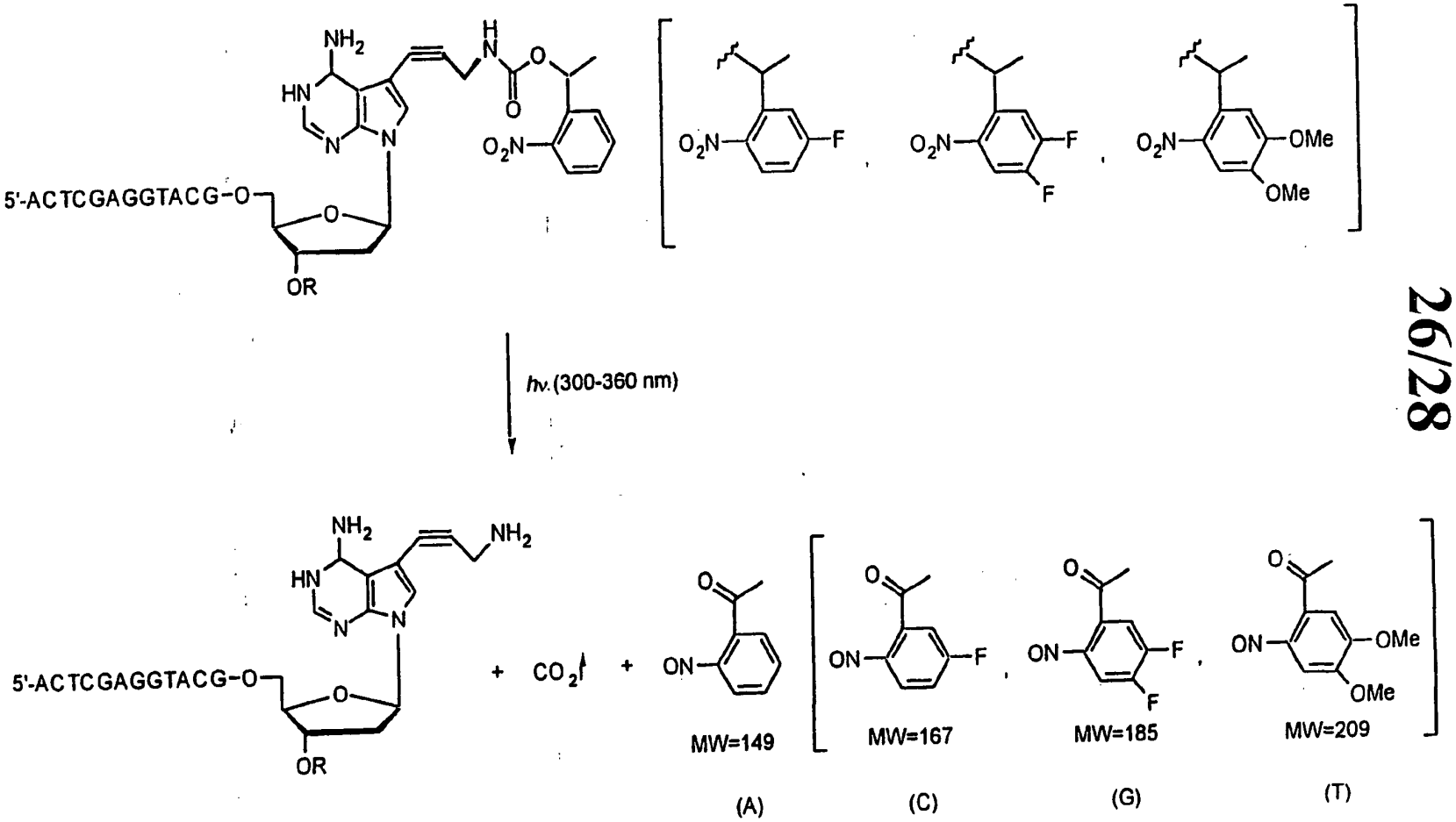

5'-ACTCGAGGTACG-O
OR
hν (300-360 nm)
+ CO2
+
MW=149
MW=167
MW=185
MW=209
(A)
(C)
(G)
(T)

FIGURE 23

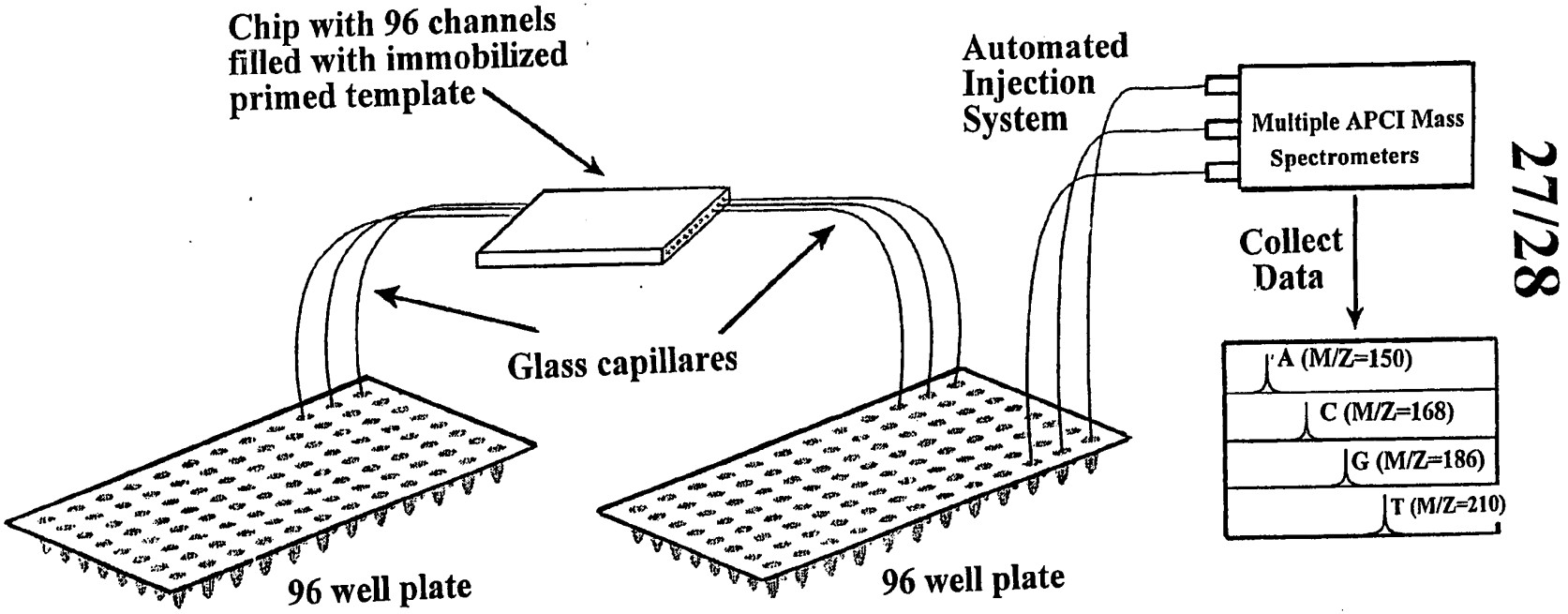
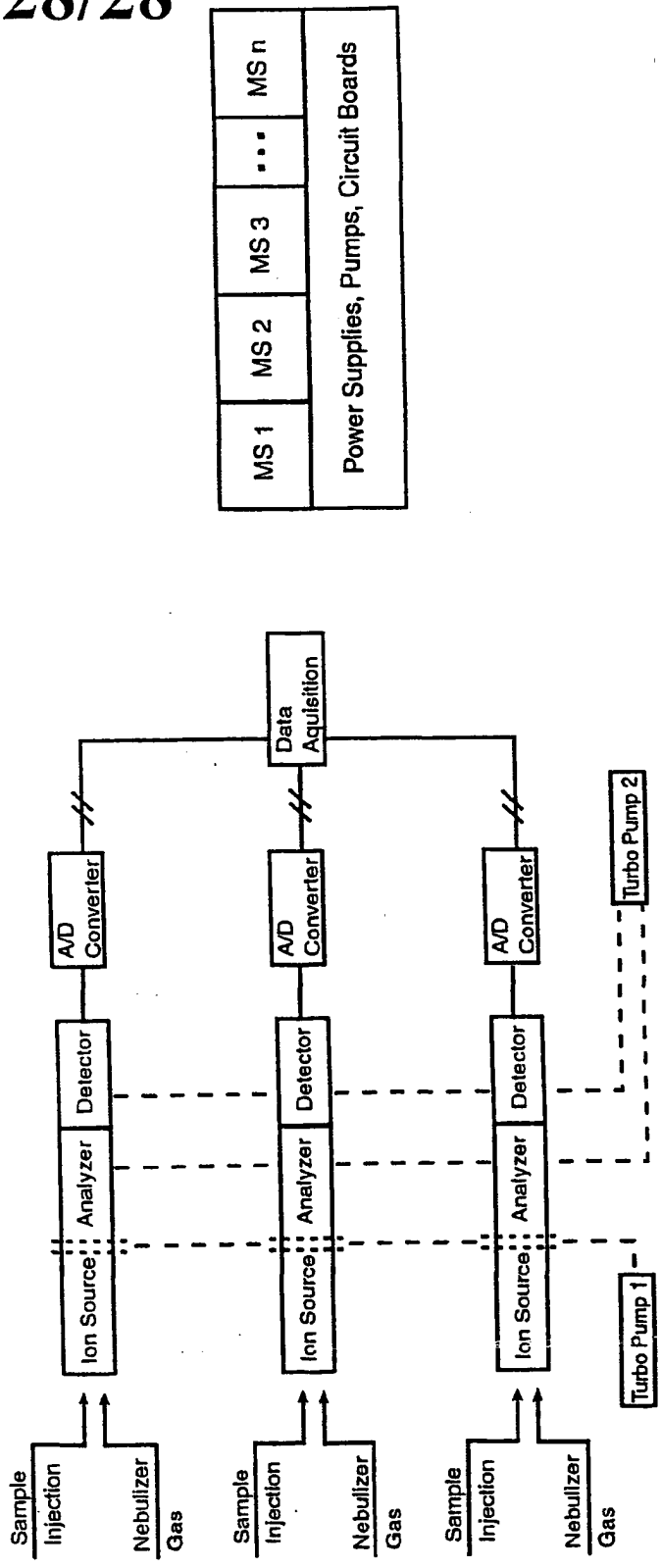

Chip with 96 channels filled with immobilized primed template
Automated Injection System
Multiple APCI Mass Spectrometers
Collect Data
A (M/Z=150)
C (M/Z=168)
G (M/Z=186)
T (M/Z=210)
Glass capillares
96 well plate
96 well plate

FIGURE 24


MS 1
MS 2
MS 3
...
MS n
Power Supplies, Pumps, Circuit Boards
Data Aquisition
A/D Converter
A/D Converter
A/D Converter
Detector
Detector
Detector
Analyzer
Analyzer
Analyzer
Ion Source
Ion Source
Ion Source
Turbo Pump 1
Turbo Pump 2
Sample Injection
Nebulizer Gas
Sample Injection
Nebulizer Gas
Sample Injection
Nebulizer Gas

SEQUENCE LISTING

<110> The Trustees Of Columbia University In The City Of

<120> Massive Parallel Method For Decoding DNA And RNA

<130> 0575/62239-B-PCT/JPW

<140> Not Yet Known
<141> 2001-10-05

<150> 09/684,670
<151> 2000-10-06

<160> 2

<170> PatentIn Ver. 2.1

<210> 1
<211> 11
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: template

<400> 1 11
acgtacgacg t

<210> 2
<211> 101
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: template

<400> 2
ttcctgcatg ggcggcatga acccgaggcc catcctcacc atcatcacac tggaagactc 60
cagtggtaat ctactgggac ggaacagctt tgaggtgcat t 101